9-21-00

A/BOX/Seq

Express Mail No.: EK916750871US
Docket No.: 789CIP2C

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
PATENT APPLICATION TRANSMITTAL UNDER 37 CFR 1.53

BOX PATENT APPLICATION
Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

Transmitted herewith for filing is the patent application of

Inventor(s): Y. Tom Tang, Chenghua Liu, Ping Zhou, Vinod Asundi, Feiyan Ren, Qing A. Zhao, Jie Zhang, Jian-Rui Wang, Tom Wehrman, Radoje T. Drmanac

Title: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

1. Type of application

☒ This is a new application for a

☒ Utility patent.

☐ Design patent.

☒ This is a continuation-in-part application of prior application no. 09/574,454 filed May 19, 2000, Attorney Docket No. 789CIP, which is a continuation-in-part application of prior application no. 09/519,705 filed March 07, 2000, Attorney Docket No. 789.

2. Application Papers Enclosed

1	Title Page	
113	Pages of Specification (excluding Claims, Abstract, Drawings & Sequence Listing)	
4	Page(s) of Claims	
1	Page(s) of Abstract	
0	Sheet(s) of Drawings (Figs. X-X)	☐ Formal ☐ Informal
102	Page(s) of Sequence Listing	

CERTIFICATION UNDER 37 CFR 1.10

I hereby certify that this Patent Application Transmittal and the documents referred to as enclosed therewith are being deposited with the United States Postal Service on **September 19, 2000,** in an envelope addressed to the Assistant Commissioner for Patents, Washington, D.C. 20231 utilizing the "Express Mail Post Office to Addressee" service of the United States Postal Service under Mailing Label No. **EK916750871US**

Leslie A. Modi

3. Oath or Declaration

☐ Enclosed

- ☐ Executed by (check all applicable boxes)
- ☐ Inventor(s)
- ☐ Legal representative of inventors(s) (37 CFR 1.42 or 1.43)
- ☐ Joint inventor or person showing a proprietary interest on behalf of inventor who refused to sign or cannot be reached
- ☐ The petition required by 37 CFR 1.47 and the statement required by 37 CFR 1.47 are enclosed. See Item 5D below for fee.

☒ Unexecuted – the undersigned attorney or agent is authorized to file this application on behalf of the applicant(s). An executed declaration will follow.

4. Additional Papers Enclosed

- ☐ Preliminary Amendment
- ☐ Information Disclosure Statement
- ☐ Declaration of Biological Deposit
- ☒ Computer readable copy of sequence listing containing nucleotide and/or amino acid sequence
- ☒ Statement Under 37 CFR § 1.821
- ☒ Paper copy of sequence listing identical to computer copy (102 pages)
- ☐ Microfiche computer program
- ☒ Verified statement claiming small entity status under 37 CFR 1.9 and 1.27
- ☐ Associate Power of Attorney
- ☐ Verified translation of a non-English patent application
- ☒ Return receipt postcard
- ☐ Other _____

5. Priority Applications Under 35 USC 119

Certified copies of applications from which priority under 35 USC 119 is claimed are listed below and

- ☐ are attached.
- ☐ will follow.

6. **Filing Fee Calculation (37 CFR 1.16)**

A. ☒ **Utility Application**

CLAIMS AS FILED – INCLUDING PRELIMINARY AMENDMENT (IF ANY)						
			SMALL ENTITY		OTHER THAN A SMALL ENTITY	
	NO. FILED	NO. EXTRA	RATE	FEE	RATE	FEE
BASIC FEE				$345.00		$690.00
TOTAL	30-20	= 10	X 9 =	$90.00	X 18 =	$0.00
INDEP.	3-3	= 0	X 39 =	$0.00	X 78 =	$0.00
☒ First Presentation of Multiple Dependent Claim			+ 130 =	$130.00	+ 260 =	$0.00
			FILING FEE:	**$565.00**	**OR**	$0.00

B. ☐ **Design Application ($155.00/$310.00)** Filing Fee: $_____

C. ☐ **Plant Application ($240.00/$480.00)** Filing Fee: $_____

D. **Other fees**

☐ Recording Assignment [Fee -- **$40.00** per assignment] $_____

☐ Other $_____

TOTAL FEES $ 565.00

7. **Method of Payments of Fees**

☐ Enclosed check

☒ Charge Deposit Account No. 501169. A duplicate copy of this transmittal is enclosed

☐ Not enclosed

8. **Deposit Account and Refund Authorization**

The Commissioner is hereby authorized to charge payment of any additional fees due or credit any overpayment to Deposit Account No. 501169. A duplicate copy of this transmittal is enclosed.

Please refund any overpayment to Hyseq, Inc. at the address below.

Please direct all future correspondence to Leslie A. Mooi at the address below.

Respectfully submitted,

Date: September 19, 2000 By: Leslie Mooi

Leslie A. Mooi
Attorney for Applicants
Registration No.: 37,047
HYSEQ, INC.
670 Almanor Avenue
Sunnyvale, CA 94085
(408) 524-8100
(408) 524-8145 (Telefacsimile)

Express Mail No.: EK916750871US
Docket No.: 789CIP2C

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s) or Patentee(s): Y. Tom Tang, Chenghua Liu, Ping Zhou, Vinod Asundi, Feiyan Ren, Qing A. Zhao, Jie Zhang, Jian-Rui Wang, Tom Wehrman, Radoje T. Drmanac

Application No. or Patent No.: Not Yet Assigned

Filed or Issued: Herewith

For: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

VERIFIED STATEMENT (DECLARATION) CLAIMING SMALL ENTITY STATUS (37 CFR § 1.9(f) AND 1.27(c)) - SMALL BUSINESS CONCERN

I hereby declare that I am

( ) The owner of the small business concern identified below:
(X) An official of the small business concern empowered to act on behalf of the concern identified below:

NAME OF CONCERN: HYSEQ, INC.
ADDRESS: 670 Almanor Avenue
Sunnyvale, CA 94085

I hereby declare that the above-identified small business concern qualifies as a small business concern as defined in 13 CFR § 121.12, and reproduced in 37 CFR § 1.9(d), for purposes of paying reduced fees under § 41(a) and (b) of Title 35, United States Code, in that the number of employees of the concern, including those of its affiliates, does not exceed 500 persons. For purposes of this statement, (1) the number of employees of the business concern is the average over the previous fiscal year of the concern of the persons employed on a full-time, part-time or temporary basis during each of the pay periods of the fiscal year, and (2) concerns are affiliates of each other when either, directly or indirectly, one concern controls or has the power to control the other, or a third party or parties controls or has the power to control both.

I hereby declare that rights under contract or law have been conveyed to, and remain with, the small business concern identified above with regard to the invention, entitled NOVEL NUCLEIC ACIDS AND POLYPEPTIDES by inventors, Y. Tom Tang, Chenghua Liu, Ping Zhou, Vinod Asundi, Feiyan Ren, Qing A. Zhao, Jie Zhang, Jian-Rui Wang, Tom Wehrman, Radoje T. Drmanac, described in

(X) The specification filed herewith.
( ) Application Serial No. [ ], filed [Date].
( ) Patent No. [ ], issued [Date].

If the rights held by the above identified small business concern are not exclusive, each individual, concern or organization having rights to the invention is listed below[1] and no rights to the invention are held by any person, other than the inventor, who could not qualify as a small business concern under 37 CFR § 1.9(c), or by any concern which would not qualify as a small business concern under 37 CFR § 1.9(d) or a nonprofit organization under 37 CFR § 1.9(e).

Full Name: _____

Address: _____

( ) Individual ( ) Small Business Concern ( ) Nonprofit Organization

I acknowledge the duty to file, in this application or patent, notification of any change in status resulting in loss of entitlement to small entity status prior to paying, or at the time of paying, the earliest of the issue fee or any maintenance fee due after the date on which status as a small entity is no longer appropriate (37 CFR § 1.28(b)).

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under § 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application, any patent issuing thereon, or any patent to which this verified statement is directed.

Name of person signing: James N. Fletcher

Title of person other than owner: Secretary

Address of person signing: HYSEQ, INC.
670 Almanor Avenue
Sunnyvale, CA 94085

Signature: _____

Date: 9/19/00

[1]NOTE: Separate verified statements are required from each named person, concern or organization having rights to the invention averring to their status as small entities. (37 CFR § 1.27)

Our Ref. No.: 789CIP2C

NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

Express Mail Label No.: EK916750871US

NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

1. CROSS REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part application of U.S. Application Serial No. 09/574,454, filed May 19, 2000, Attorney Docket No. 789CIP, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/519,705, filed March 07, 2000, Attorney Docket No. 789, both of which are incorporated herein by reference in their entirety.

2. BACKGROUND OF THE INVENTION

2.1 TECHNICAL FIELD

The present invention provides novel polynucleotides and proteins encoded by such polynucleotides, along with uses for these polynucleotides and proteins, for example in therapeutic, diagnostic and research methods.

2.2 BACKGROUND

Technology aimed at the discovery of protein factors (including e.g., cytokines, such as lymphokines, interferons, CSFs, chemokines, and interleukins) has matured rapidly over the past decade. The now routine hybridization cloning and expression cloning techniques clone novel polynucleotides "directly" in the sense that they rely on information directly related to the discovered protein (i.e., partial DNA/amino acid sequence of the protein in the case of hybridization cloning; activity of the protein in the case of expression cloning). More recent "indirect" cloning techniques such as signal sequence cloning, which isolates DNA sequences based on the presence of a now well-recognized secretory leader sequence motif, as well as various PCR-based or low stringency hybridization-based cloning techniques, have advanced the state of the art by making available large numbers of DNA/amino acid sequences for proteins that are known to have biological activity, for example, by virtue of their secreted nature in the case of leader sequence cloning, by virtue of their cell or tissue source in the case of

PCR-based techniques, or by virtue of structural similarity to other genes of known biological activity.

Identified polynucleotide and polypeptide sequences have numerous applications in, for example, diagnostics, forensics, gene mapping; identification of mutations responsible for genetic disorders or other traits, to assess biodiversity, and to produce many other types of data and products dependent on DNA and amino acid sequences.

3. SUMMARY OF THE INVENTION

The compositions of the present invention include novel isolated polypeptides, novel isolated polynucleotides encoding such polypeptides, including recombinant DNA molecules, cloned genes or degenerate variants thereof, especially naturally occurring variants such as allelic variants, antisense polynucleotide molecules, and antibodies that specifically recognize one or more epitopes present on such polypeptides, as well as hybridomas producing such antibodies.

The compositions of the present invention additionally include vectors, including expression vectors, containing the polynucleotides of the invention, cells genetically engineered to contain such polynucleotides and cells genetically engineered to express such polynucleotides.

The present invention relates to a collection or library of at least one novel nucleic acid sequence assembled from expressed sequence tags (ESTs) isolated mainly by sequencing by hybridization (SBH), and in some cases, sequences obtained from one or more public databases. The invention relates also to the proteins encoded by such polynucleotides, along with therapeutic, diagnostic and research utilities for these polynucleotides and proteins. These nucleic acid sequences are designated as SEQ ID NO: 1 – 35 and are provided in the Sequence Listing. In the nucleic acids provided in the Sequence Listing, A is adenosine; C is cytosine; G is guanosine; T is thymine; and N is any of the four bases. In the amino acids provided in the Sequence Listing, * corresponds to the stop codon.

The nucleic acid sequences of the present invention also include, nucleic acid sequences that hybridize to the complement of SEQ ID NO: 1 – 35 under stringent hybridization conditions; nucleic acid sequences which are allelic variants or species

homologues of any of the nucleic acid sequences recited above, or nucleic acid sequences that encode a peptide comprising a specific domain or truncation of the peptides encoded by SEQ ID NO: 1 – 35. A polynucleotide comprising a nucleotide sequence having at least 90% identity to an identifying sequence of SEQ ID NO: 1 – 35 or a degenerate variant or fragment thereof. The identifying sequence can be 100 base pairs in length.

The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NO: 1 – 35. The sequence information can be a segment of any one of SEQ ID NO: 1 – 35 that uniquely identifies or represents the sequence information of SEQ ID NO: 1 – 35.

A collection as used in this application can be a collection of only one polynucleotide. The collection of sequence information or identifying information of each sequence can be provided on a nucleic acid array. In one embodiment, segments of sequence information is provided on a nucleic acid array to detect the polynucleotide that contains the segment. The array can be designed to detect full-match or mismatch to the polynucleotide that contains the segment. The collection can also be provided in a computer-readable format.

This invention also includes the reverse or direct complement of any of the nucleic acid sequences recited above; cloning or expression vectors containing the nucleic acid sequences; and host cells or organisms transformed with these expression vectors. Nucleic acid sequences (or their reverse or direct complements) according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology, such as use as hybridization probes, use as primers for PCR, use in an array, use in computer-readable media, use in sequencing full-length genes, use for chromosome and gene mapping, use in the recombinant production of protein, and use in the generation of anti-sense DNA or RNA, their chemical analogs and the like.

In a preferred embodiment, the nucleic acid sequences of SEQ ID NO: 1-35 or novel segments or parts of the nucleic acids of the invention are used as primers in expression assays that are well known in the art. In a particularly preferred embodiment, the nucleic acid sequences of SEQ ID NO: 1-35 or novel segments or parts of the nucleic acids provided herein are used in diagnostics for identifying expressed genes or, as well known in

the art and exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The isolated polynucleotides of the invention include, but are not limited to, a polynucleotide comprising any one of the nucleotide sequences set forth in the SEQ ID NO: 1–35; a polynucleotide comprising any of the full length protein coding sequences of the SEQ ID NO: 1–35; and a polynucleotide comprising any of the nucleotide sequences of the mature protein coding sequences of the SEQ ID NO: 1–35. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent hybridization conditions to (a) the complement of any one of the nucleotide sequences set forth in the SEQ ID NO: 1–35; (b) a nucleotide sequence encoding any one of the amino acid sequences set forth in the Sequence Listing; (c) a polynucleotide which is an allelic variant of any polynucleotides recited above; (d) a polynucleotide which encodes a species homolog (e.g. orthologs) of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of any of the polypeptides comprising an amino acid sequence set forth in the Sequence Listing.

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising any of the amino acid sequences set forth in the Sequence Listing; or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides with biological activity that are encoded by (a) any of the polynucleotides having a nucleotide sequence set forth in the SEQ ID NO: 1-35; or (b) polynucleotides that hybridize to the complement of the polynucleotides of (a) under stringent hybridization conditions. Biologically or immunologically active variants of any of the polypeptide sequences in the Sequence Listing, and "substantial equivalents" thereof (e.g., with at least about 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% amino acid sequence identity) that preferably retain biological activity are also contemplated. The polypeptides of the invention may be wholly or partially chemically synthesized but are preferably produced by recombinant means using the genetically engineered cells (e.g. host cells) of the invention.

The invention also provides compositions comprising a polypeptide of the invention. Polypeptide compositions of the invention may further comprise an acceptable carrier, such as a hydrophilic, e.g., pharmaceutically acceptable, carrier.

The invention also provides host cells transformed or transfected with a polynucleotide of the invention.

The invention also relates to methods for producing a polypeptide of the invention comprising growing a culture of the host cells of the invention in a suitable culture medium under conditions permitting expression of the desired polypeptide, and purifying the polypeptide from the culture or from the host cells. Preferred embodiments include those in which the protein produced by such process is a mature form of the protein.

Polynucleotides according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology. These techniques include use as hybridization probes, use as oligomers, or primers, for PCR, use for chromosome and gene mapping, use in the recombinant production of protein, and use in generation of anti-sense DNA or RNA, their chemical analogs and the like. For example, when the expression of an mRNA is largely restricted to a particular cell or tissue type, polynucleotides of the invention can be used as hybridization probes to detect the presence of the particular cell or tissue mRNA in a sample using, *e.g., in situ* hybridization.

In other exemplary embodiments, the polynucleotides are used in diagnostics as expressed sequence tags for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The polypeptides according to the invention can be used in a variety of conventional procedures and methods that are currently applied to other proteins. For example, a polypeptide of the invention can be used to generate an antibody that specifically binds the polypeptide. Such antibodies, particularly monoclonal antibodies, are useful for detecting or quantitating the polypeptide in tissue. The polypeptides of the invention can also be used as molecular weight markers, and as a food supplement.

Methods are also provided for preventing, treating, or ameliorating a medical condition which comprises the step of administering to a mammalian subject a therapeutically effective amount of a composition comprising a polypeptide of the present invention and a pharmaceutically acceptable carrier.

In particular, the polypeptides and polynucleotides of the invention can be utilized, for example, in methods for the prevention and/or treatment of disorders involving aberrant protein expression or biological activity.

The present invention further relates to methods for detecting the presence of the polynucleotides or polypeptides of the invention in a sample. Such methods can, for example, be utilized as part of prognostic and diagnostic evaluation of disorders as recited herein and for the identification of subjects exhibiting a predisposition to such conditions. The invention provides a method for detecting the polynucleotides of the invention in a sample, comprising contacting the sample with a compound that binds to and forms a complex with the polynucleotide of interest for a period sufficient to form the complex and under conditions sufficient to form a complex and detecting the complex such that if a complex is detected, the polynucleotide of interest is detected. The invention also provides a method for detecting the polypeptides of the invention in a sample comprising contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex and detecting the formation of the complex such that if a complex is formed, the polypeptide is detected.

The invention also provides kits comprising polynucleotide probes and/or monoclonal antibodies, and optionally quantitative standards, for carrying out methods of the invention. Furthermore, the invention provides methods for evaluating the efficacy of drugs, and monitoring the progress of patients, involved in clinical trials for the treatment of disorders as recited above.

The invention also provides methods for the identification of compounds that modulate (i.e., increase or decrease) the expression or activity of the polynucleotides and/or polypeptides of the invention. Such methods can be utilized, for example, for the identification of compounds that can ameliorate symptoms of disorders as recited herein. Such methods can include, but are not limited to, assays for identifying compounds and other substances that interact with (*e.g.*, bind to) the polypeptides of the invention. The invention provides a method for identifying a compound that binds to the polypeptides of the invention comprising contacting the compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the

complex drives expression of a reporter gene sequence in the cell; and detecting the complex by detecting the reporter gene sequence expression such that if expression of the reporter gene is detected the compound the binds to a polypeptide of the invention is identified.

The methods of the invention also provides methods for treatment which involve the administration of the polynucleotides or polypeptides of the invention to individuals exhibiting symptoms or tendencies. In addition, the invention encompasses methods for treating diseases or disorders as recited herein comprising administering compounds and other substances that modulate the overall activity of the target gene products. Compounds and other substances can effect such modulation either on the level of target gene/protein expression or target protein activity.

The polypeptides of the present invention and the polynucleotides encoding them are also useful for the same functions known to one of skill in the art as the polypeptides and polynucleotides to which they have homology (set forth in Table 1); for which they have a signature region (as set forth in Table 3); or for which they have homology to a gene family (as set forth in Table 4). If no homology is set forth for a sequence, then the polypeptides and polynucleotides of the present invention are useful for a variety of applications, as described herein, including use in arrays for detection.

4. DETAILED DESCRIPTION OF THE INVENTION

4.1 DEFINITIONS

It must be noted that as used herein and in the appended claims, the singular forms "a", "an" and "the" include plural references unless the context clearly dictates otherwise.

The term "active" refers to those forms of the polypeptide which retain the biologic and/or immunologic activities of any naturally occurring polypeptide. According to the invention, the terms "biologically active" or "biological activity" refer to a protein or peptide having structural, regulatory or biochemical functions of a naturally occurring molecule. Likewise "immunologically active" or "immunological activity" refers to the

capability of the natural, recombinant or synthetic polypeptide to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The term "activated cells" as used in this application are those cells which are engaged in extracellular or intracellular membrane trafficking, including the export of secretory or enzymatic molecules as part of a normal or disease process.

The terms "complementary" or "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence 5'-AGT-3' binds to the complementary sequence 3'-TCA-5'. Complementarity between two single-stranded molecules may be "partial" such that only some of the nucleic acids bind or it may be "complete" such that total complementarity exists between the single stranded molecules. The degree of complementarity between the nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands.

The term "embryonic stem cells (ES)" refers to a cell that can give rise to many differentiated cell types in an embryo or an adult, including the germ cells. The term "germ line stem cells (GSCs)" refers to stem cells derived from primordial stem cells that provide a steady and continuous source of germ cells for the production of gametes. The term "primordial germ cells (PGCs)" refers to a small population of cells set aside from other cell lineages particularly from the yolk sac, mesenteries, or gonadal ridges during embryogenesis that have the potential to differentiate into germ cells and other cells. PGCs are the source from which GSCs and ES cells are derived The PGCs, the GSCs and the ES cells are capable of self-renewal. Thus these cells not only populate the germ line and give rise to a plurality of terminally differentiated cells that comprise the adult specialized organs, but are able to regenerate themselves.

The term "expression modulating fragment," EMF, means a series of nucleotides which modulates the expression of an operably linked ORF or another EMF.

As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs are nucleic acid fragments which induce the expression of an operably linked ORF in response to a specific regulatory factor or physiological event.

The terms "nucleotide sequence" or "nucleic acid" or "polynucleotide" or "oligonculeotide" are used interchangeably and refer to a heteropolymer of nucleotides or the sequence of these nucleotides. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA) or to any DNA-like or RNA-like material. It is contemplated that where the polynucleotide is RNA, the T (thymine) in the sequences provided herein is substituted with U (uracil). Generally, nucleic acid segments provided by this invention may be assembled from fragments of the genome and short oligonucleotide linkers, or from a series of oligonucleotides, or from individual nucleotides, to provide a synthetic nucleic acid which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon, or a eukaryotic gene.

The terms "oligonucleotide fragment" or a "polynucleotide fragment", "portion," or "segment" or "probe" or "primer" are used interchangeable and refer to a sequence of nucleotide residues which are at least about 5 nucleotides, more preferably at least about 7 nucleotides, more preferably at least about 9 nucleotides, more preferably at least about 11 nucleotides and most preferably at least about 17 nucleotides. The fragment is preferably less than about 500 nucleotides, preferably less than about 200 nucleotides, more preferably less than about 100 nucleotides, more preferably less than about 50 nucleotides and most preferably less than 30 nucleotides. Preferably the probe is from about 6 nucleotides to about 200 nucleotides, preferably from about 15 to about 50 nucleotides, more preferably from about 17 to 30 nucleotides and most preferably from about 20 to 25 nucleotides. Preferably the fragments can be used in polymerase chain reaction (PCR), various hybridization procedures or microarray procedures to identify or amplify identical or related parts of mRNA or DNA molecules. A fragment or segment may uniquely identify each polynucleotide sequence of the present invention. Preferably the fragment comprises a sequence substantially similar to any one of SEQ ID NOs:1-35.

Probes may, for example, be used to determine whether specific mRNA molecules are present in a cell or tissue or to isolate similar nucleic acid sequences from chromosomal DNA as described by Walsh et al. (Walsh, P.S. et al., 1992, PCR Methods Appl 1:241-250). They may be labeled by nick translation, Klenow fill-in reaction, PCR,

or other methods well known in the art. Probes of the present invention, their preparation and/or labeling are elaborated in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY; or Ausubel, F.M. et al., 1989, Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, both of which are incorporated herein by reference in their entirety.

The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NOs: 1-35. The sequence information can be a segment of any one of SEQ ID NOs: 1-35 that uniquely identifies or represents the sequence information of that sequence of SEQ ID NO: 1-35. One such segment can be a twenty-mer nucleic acid sequence because the probability that a twenty-mer is fully matched in the human genome is 1 in 300. In the human genome, there are three billion base pairs in one set of chromosomes. Because 4^{20} possible twenty-mers exist, there are 300 times more twenty-mers than there are base pairs in a set of human chromosome. Using the same analysis, the probability for a seventeen-mer to be fully matched in the human genome is approximately 1 in 5. When these segments are used in arrays for expression studies, fifteen-mer segments can be used. The probability that the fifteen-mer is fully matched in the expressed sequences is also approximately one in five because expressed sequences comprise less than approximately 5% of the entire genome sequence.

Similarly, when using sequence information for detecting a single mismatch, a segment can be a twenty-five mer. The probability that the twenty-five mer would appear in a human genome with a single mismatch is calculated by multiplying the probability for a full match ($1 \div 4^{25}$) times the increased probability for mismatch at each nucleotide position (3 x 25). The probability that an eighteen mer with a single mismatch can be detected in an array for expression studies is approximately one in five. The probability that a twenty-mer with a single mismatch can be detected in a human genome is approximately one in five.

The term "open reading frame," ORF, means a series of nucleotide triplets coding for amino acids without any termination codons and is a sequence translatable into protein.

The terms "operably linked" or "operably associated" refer to functionally related nucleic acid sequences. For example, a promoter is operably associated or operably

linked with a coding sequence if the promoter controls the transcription of the coding sequence. While operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements e.g. repressor genes are not contiguously linked to the coding sequence but still control transcription/translation of the coding sequence.

The term "pluripotent" refers to the capability of a cell to differentiate into a number of differentiated cell types that are present in an adult organism. A pluripotent cell is restricted in its differentiation capability in comparison to a totipotent cell.

The terms "polypeptide" or "peptide" or "amino acid sequence" refer to an oligopeptide, peptide, polypeptide or protein sequence or fragment thereof and to naturally occurring or synthetic molecules. A polypeptide "fragment," "portion," or "segment" is a stretch of amino acid residues of at least about 5 amino acids, preferably at least about 7 amino acids, more preferably at least about 9 amino acids and most preferably at least about 17 or more amino acids. The peptide preferably is not greater than about 200 amino acids, more preferably less than 150 amino acids and most preferably less than 100 amino acids. Preferably the peptide is from about 5 to about 200 amino acids. To be active, any polypeptide must have sufficient length to display biological and/or immunological activity.

The term "naturally occurring polypeptide" refers to polypeptides produced by cells that have not been genetically engineered and specifically contemplates various polypeptides arising from post-translational modifications of the polypeptide including, but not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation.

The term "translated protein coding portion" means a sequence which encodes for the full length protein which may include any leader sequence or any processing sequence.

The term "mature protein coding sequence" means a sequence which encodes a peptide or protein without a signal or leader sequence. The peptide may have been produced by processing in the cell which removes any leader/signal sequence. The peptide may be produced synthetically or the protein may have been produced using a polynucleotide only encoding for the mature protein coding sequence.

The term "derivative" refers to polypeptides chemically modified by such techniques as ubiquitination, labeling (e.g., with radionuclides or various enzymes), covalent polymer attachment such as pegylation (derivatization with polyethylene glycol) and insertion or substitution by chemical synthesis of amino acids such as ornithine, which do not normally occur in human proteins.

The term "variant"(or "analog") refers to any polypeptide differing from naturally occurring polypeptides by amino acid insertions, deletions, and substitutions, created using, *e g.*, recombinant DNA techniques. Guidance in determining which amino acid residues may be replaced, added or deleted without abolishing activities of interest, may be found by comparing the sequence of the particular polypeptide with that of homologous peptides and minimizing the number of amino acid sequence changes made in regions of high homology (conserved regions) or by replacing amino acids with consensus sequence.

Alternatively, recombinant variants encoding these same or similar polypeptides may be synthesized or selected by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce various restriction sites, may be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic system. Mutations in the polynucleotide sequence may be reflected in the polypeptide or domains of other peptides added to the polypeptide to modify the properties of any part of the polypeptide, to change characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate.

Preferably, amino acid "substitutions" are the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, *i.e.*, conservative amino acid replacements. "Conservative" amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino

acids include aspartic acid and glutamic acid. "Insertions" or "deletions" are preferably in the range of about 1 to 20 amino acids, more preferably 1 to 10 amino acids. The variation allowed may be experimentally determined by systematically making insertions, deletions, or substitutions of amino acids in a polypeptide molecule using recombinant DNA techniques and assaying the resulting recombinant variants for activity.

Alternatively, where alteration of function is desired, insertions, deletions or non-conservative alterations can be engineered to produce altered polypeptides. Such alterations can, for example, alter one or more of the biological functions or biochemical characteristics of the polypeptides of the invention. For example, such alterations may change polypeptide characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate. Further, such alterations can be selected so as to generate polypeptides that are better suited for expression, scale up and the like in the host cells chosen for expression. For example, cysteine residues can be deleted or substituted with another amino acid residue in order to eliminate disulfide bridges.

The terms "purified" or "substantially purified" as used herein denotes that the indicated nucleic acid or polypeptide is present in the substantial absence of other biological macromolecules, *e.g.*, polynucleotides, proteins, and the like. In one embodiment, the polynucleotide or polypeptide is purified such that it constitutes at least 95% by weight, more preferably at least 99% by weight, of the indicated biological macromolecules present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 1000 daltons, can be present).

The term "isolated" as used herein refers to a nucleic acid or polypeptide separated from at least one other component (e.g., nucleic acid or polypeptide) present with the nucleic acid or polypeptide in its natural source. In one embodiment, the nucleic acid or polypeptide is found in the presence of (if anything) only a solvent, buffer, ion, or other component normally present in a solution of the same. The terms "isolated" and "purified" do not encompass nucleic acids or polypeptides present in their natural source.

The term "recombinant," when used herein to refer to a polypeptide or protein, means that a polypeptide or protein is derived from recombinant (*e.g.*, microbial, insect, or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (e.g., yeast) expression systems. As a product,

"recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, *e.g., E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern in general different from those expressed in mammalian cells.

The term "recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. An expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers, (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an amino terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

The term "recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extrachromosomally. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed. This term also means host cells which have stably integrated a recombinant genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers. Recombinant expression systems as defined herein will express polypeptides or proteins endogenous to the cell upon induction of the regulatory elements linked to the endogenous DNA segment or gene to be expressed. The cells can be prokaryotic or eukaryotic.

The term "secreted" includes a protein that is transported across or through a membrane, including transport as a result of signal sequences in its amino acid sequence when it is expressed in a suitable host cell. "Secreted" proteins include without limitation

proteins secreted wholly (*e.g.*, soluble proteins) or partially (*e.g.*, receptors) from the cell in which they are expressed. "Secreted" proteins also include without limitation proteins that are transported across the membrane of the endoplasmic reticulum. "Secreted" proteins are also intended to include proteins containing non-typical signal sequences (e.g. Interleukin-1 Beta, see Krasney, P.A. and Young, P.R. (1992) Cytokine 4(2):134 -143) and factors released from damaged cells (e.g. Interleukin-1 Receptor Antagonist, see Arend, W.P. et. al. (1998) Annu. Rev. Immunol. 16:27-55)

Where desired, an expression vector may be designed to contain a "signal or leader sequence" which will direct the polypeptide through the membrane of a cell. Such a sequence may be naturally present on the polypeptides of the present invention or provided from heterologous protein sources by recombinant DNA techniques.

The term "stringent" is used to refer to conditions that are commonly understood in the art as stringent. Stringent conditions can include highly stringent conditions (i.e., hybridization to filter-bound DNA in 0.5 M $NaHPO_4$, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1X SSC/0.1% SDS at 68°C), and moderately stringent conditions (i.e., washing in 0.2X SSC/0.1% SDS at 42°C). Other exemplary hybridization conditions are described herein in the examples.

In instances of hybridization of deoxyoligonucleotides, additional exemplary stringent hybridization conditions include washing in 6X SSC/0.05% sodium pyrophosphate at 37°C (for 14-base oligonucleotides), 48°C (for 17-base oligos), 55°C (for 20-base oligonucleotides), and 60°C (for 23-base oligonucleotides).

As used herein, "substantially equivalent" can refer both to nucleotide and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between the reference and subject sequences. Typically, such a substantially equivalent sequence varies from one of those listed herein by no more than about 35% (*i.e.*, the number of individual residue substitutions, additions, and/or deletions in a substantially equivalent sequence, as compared to the corresponding reference sequence, divided by the total number of residues in the substantially equivalent sequence is about 0.35 or less). Such a sequence is said to have 65% sequence identity to the listed sequence. In one embodiment, a substantially

equivalent, *e.g.*, mutant, sequence of the invention varies from a listed sequence by no more than 30% (70% sequence identity); in a variation of this embodiment, by no more than 25% (75% sequence identity); and in a further variation of this embodiment, by no more than 20% (80% sequence identity) and in a further variation of this embodiment, by no more than 10% (90% sequence identity) and in a further variation of this embodiment, by no more that 5% (95% sequence identity). Substantially equivalent, *e.g.*, mutant, amino acid sequences according to the invention preferably have at least 80% sequence identity with a listed amino acid sequence, more preferably at least 90% sequence identity. Substantially equivalent nucleotide sequences of the invention can have lower percent sequence identities, taking into account, for example, the redundancy or degeneracy of the genetic code. Preferably, nucleotide sequence has at least about 65% identity, more preferably at least about 75% identity, and most preferably at least about 95% identity. For the purposes of the present invention, sequences having substantially equivalent biological activity and substantially equivalent expression characteristics are considered substantially equivalent. For the purposes of determining equivalence, truncation of the mature sequence (*e.g.*, via a mutation which creates a spurious stop codon) should be disregarded. Sequence identity may be determined, e.g., using the Jotun Hein method (Hein, J. (1990) Methods Enzymol. 183:626-645). Identity between sequences can also be determined by other methods known in the art, e.g. by varying hybridization conditions.

The term "totipotent" refers to the capability of a cell to differentiate into all of the cell types of an adult organism.

The term "transformation" means introducing DNA into a suitable host cell so that the DNA is replicable, either as an extrachromosomal element, or by chromosomal integration. The term "transfection" refers to the taking up of an expression vector by a suitable host cell, whether or not any coding sequences are in fact expressed. The term "infection" refers to the introduction of nucleic acids into a suitable host cell by use of a virus or viral vector.

As used herein, an "uptake modulating fragment," UMF, means a series of nucleotides which mediate the uptake of a linked DNA fragment into a cell. UMFs can be readily identified using known UMFs as a target sequence or target motif with the

computer-based systems described below. The presence and activity of a UMF can be confirmed by attaching the suspected UMF to a marker sequence. The resulting nucleic acid molecule is then incubated with an appropriate host under appropriate conditions and the uptake of the marker sequence is determined. As described above, a UMF will increase the frequency of uptake of a linked marker sequence.

Each of the above terms is meant to encompass all that is described for each, unless the context dictates otherwise.

4.2 NUCLEIC ACIDS OF THE INVENTION

Nucleotide sequences of the invention are set forth in the Sequence Listing.

The isolated polynucleotides of the invention include a polynucleotide comprising the nucleotide sequences of the SEQ ID NO: 1 – 35; a polynucleotide encoding any one of the peptide sequences of SEQ ID NO:1 – 35; and a polynucleotide comprising the nucleotide sequence encoding the mature protein coding sequence of the polynucleotides of any one of SEQ ID NO: 1 - 35. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent conditions to (a) the complement of any of the nucleotides sequences of the SEQ ID NO: 1 – 35; (b) nucleotide sequences encoding any one of the amino acid sequences set forth in the Sequence Listing; (c) a polynucleotide which is an allelic variant of any polynucleotide recited above; (d) a polynucleotide which encodes a species homolog of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of the polypeptides of SEQ ID NO: 1- 35. Domains of interest may depend on the nature of the encoded polypeptide; e.g., domains in receptor-like polypeptides include ligand-binding, extracellular, transmembrane, or cytoplasmic domains, or combinations thereof; domains in immunoglobulin-like proteins include the variable immunoglobulin-like domains; domains in enzyme-like polypeptides include catalytic and substrate binding domains; and domains in ligand polypeptides include receptor-binding domains.

The polynucleotides of the invention include naturally occurring or wholly or partially synthetic DNA, e.g., cDNA and genomic DNA, and RNA, e.g., mRNA. The

polynucleotides may include all of the coding region of the cDNA or may represent a portion of the coding region of the cDNA.

The present invention also provides genes corresponding to the cDNA sequences disclosed herein. The corresponding genes can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include the preparation of probes or primers from the disclosed sequence information for identification and/or amplification of genes in appropriate genomic libraries or other sources of genomic materials. Further 5' and 3' sequence can be obtained using methods known in the art. For example, full length cDNA or genomic DNA that corresponds to any of the polynucleotides of the SEQ ID NO: 1 – 35 can be obtained by screening appropriate cDNA or genomic DNA libraries under suitable hybridization conditions using any of the polynucleotides of the SEQ ID NO: 1 - 35 or a portion thereof as a probe. Alternatively, the polynucleotides of the SEQ ID NO: 1 - 35 may be used as the basis for suitable primer(s) that allow identification and/or amplification of genes in appropriate genomic DNA or cDNA libraries.

The nucleic acid sequences of the invention can be assembled from ESTs and sequences (including cDNA and genomic sequences) obtained from one or more public databases, such as dbEST, gbpri, and UniGene. The EST sequences can provide identifying sequence information, representative fragment or segment information, or novel segment information for the full-length gene.

The polynucleotides of the invention also provide polynucleotides including nucleotide sequences that are substantially equivalent to the polynucleotides recited above. Polynucleotides according to the invention can have, *e.g.*, at least about 65%, at least about 70%, at least about 75%, at least about 80%, more typically at least about 90%, and even more typically at least about 95%, sequence identity to a polynucleotide recited above.

Included within the scope of the nucleic acid sequences of the invention are nucleic acid sequence fragments that hybridize under stringent conditions to any of the nucleotide sequences of the SEQ ID NO: 1 - 35, or complements thereof, which fragment is greater than about 5 nucleotides, preferably 7 nucleotides, more preferably greater than 9 nucleotides and most preferably greater than 17 nucleotides. Fragments of, e.g. 15, 17, or 20 nucleotides or more that are selective for (i.e. specifically hybridize to any one of

the polynucleotides of the invention) are contemplated. Probes capable of specifically hybridizing to a polynucleotide can differentiate polynucleotide sequences of the invention from other polynucleotide sequences in the same family of genes or can differentiate human genes from genes of other species, and are preferably based on unique nucleotide sequences.

The sequences falling within the scope of the present invention are not limited to these specific sequences, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequence provided in SEQ ID NO: 1 - 35, a representative fragment thereof, or a nucleotide sequence at least 90% identical, preferably 95% identical, to SEQ ID NOs: 1 - 35 with a sequence from another isolate of the same species. Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the specific ORFs disclosed herein. In other words, in the coding region of an ORF, substitution of one codon for another codon that encodes the same amino acid is expressly contemplated.

The nearest neighbor or homology result for the nucleic acids of the present invention, including SEQ ID NOs: 1 - 35, can be obtained by searching a database using an algorithm or a program. Preferably, a BLAST which stands for Basic Local Alignment Search Tool is used to search for local sequence alignments (Altshul, S.F. J Mol. Evol. 36 290-300 (1993) and Altschul S.F. et al. J. Mol. Biol. 21:403-410 (1990)). Alternatively a FASTA version 3 search against Genpept, using Fastxy algorithm.

Species homologs (or orthologs) of the disclosed polynucleotides and proteins are also provided by the present invention. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source from the desired species.

The invention also encompasses allelic variants of the disclosed polynucleotides or proteins; that is, naturally-occurring alternative forms of the isolated polynucleotide which also encode proteins which are identical, homologous or related to that encoded by the polynucleotides.

The nucleic acid sequences of the invention are further directed to sequences which encode variants of the described nucleic acids. These amino acid sequence variants may be prepared by methods known in the art by introducing appropriate

nucleotide changes into a native or variant polynucleotide. There are two variables in the construction of amino acid sequence variants: the location of the mutation and the nature of the mutation. Nucleic acids encoding the amino acid sequence variants are preferably constructed by mutating the polynucleotide to encode an amino acid sequence that does not occur in nature. These nucleic acid alterations can be made at sites that differ in the nucleic acids from different species (variable positions) or in highly conserved regions (constant regions). Sites at such locations will typically be modified in series, *e.g.*, by substituting first with conservative choices (*e.g.*, hydrophobic amino acid to a different hydrophobic amino acid) and then with more distant choices (*e.g.*, hydrophobic amino acid to a charged amino acid), and then deletions or insertions may be made at the target site. Amino acid sequence deletions generally range from about 1 to 30 residues, preferably about 1 to 10 residues, and are typically contiguous. Amino acid insertions include amino- and/or carboxyl-terminal fusions ranging in length from one to one hundred or more residues, as well as intrasequence insertions of single or multiple amino acid residues. Intrasequence insertions may range generally from about 1 to 10 amino residues, preferably from 1 to 5 residues. Examples of terminal insertions include the heterologous signal sequences necessary for secretion or for intracellular targeting in different host cells and sequences such as FLAG or poly-histidine sequences useful for purifying the expressed protein.

In a preferred method, polynucleotides encoding the novel amino acid sequences are changed via site-directed mutagenesis. This method uses oligonucleotide sequences to alter a polynucleotide to encode the desired amino acid variant, as well as sufficient adjacent nucleotides on both sides of the changed amino acid to form a stable duplex on either side of the site of being changed. In general, the techniques of site-directed mutagenesis are well known to those of skill in the art and this technique is exemplified by publications such as, Edelman et al., *DNA* 2:183 (1983). A versatile and efficient method for producing site-specific changes in a polynucleotide sequence was published by Zoller and Smith, *Nucleic Acids Res.* 10:6487-6500 (1982). PCR may also be used to create amino acid sequence variants of the novel nucleic acids. When small amounts of template DNA are used as starting material, primer(s) that differs slightly in sequence from the corresponding region in the template DNA can generate the desired amino acid

variant. PCR amplification results in a population of product DNA fragments that differ from the polynucleotide template encoding the polypeptide at the position specified by the primer. The product DNA fragments replace the corresponding region in the plasmid and this gives a polynucleotide encoding the desired amino acid variant.

A further technique for generating amino acid variants is the cassette mutagenesis technique described in Wells et al., *Gene* 34:315 (1985); and other mutagenesis techniques well known in the art, such as, for example, the techniques in Sambrook et al., supra, and *Current Protocols in Molecular Biology*, Ausubel et al. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be used in the practice of the invention for the cloning and expression of these novel nucleic acids. Such DNA sequences include those which are capable of hybridizing to the appropriate novel nucleic acid sequence under stringent conditions.

Polynucleotides encoding preferred polypeptide truncations of the invention can be used to generate polynucleotides encoding chimeric or fusion proteins comprising one or more domains of the invention and heterologous protein sequences.

The polynucleotides of the invention additionally include the complement of any of the polynucleotides recited above. The polynucleotide can be DNA (genomic, cDNA, amplified, or synthetic) or RNA. Methods and algorithms for obtaining such polynucleotides are well known to those of skill in the art and can include, for example, methods for determining hybridization conditions that can routinely isolate polynucleotides of the desired sequence identities.

In accordance with the invention, polynucleotide sequences comprising the mature protein coding sequences corresponding to any one of SEQ ID NO: 1-35, or functional equivalents thereof, may be used to generate recombinant DNA molecules that direct the expression of that nucleic acid, or a functional equivalent thereof, in appropriate host cells. Also included are the cDNA inserts of any of the clones identified herein.

A polynucleotide according to the invention can be joined to any of a variety of other nucleotide sequences by well-established recombinant DNA techniques (see Sambrook J et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor

Laboratory, NY). Useful nucleotide sequences for joining to polynucleotides include an assortment of vectors, e.g., plasmids, cosmids, lambda phage derivatives, phagemids, and the like, that are well known in the art. Accordingly, the invention also provides a vector including a polynucleotide of the invention and a host cell containing the polynucleotide. In general, the vector contains an origin of replication functional in at least one organism, convenient restriction endonuclease sites, and a selectable marker for the host cell. Vectors according to the invention include expression vectors, replication vectors, probe generation vectors, and sequencing vectors. A host cell according to the invention can be a prokaryotic or eukaryotic cell and can be a unicellular organism or part of a multicellular organism.

The present invention further provides recombinant constructs comprising a nucleic acid having any of the nucleotide sequences of the SEQ ID NOs: 1 - 35 or a fragment thereof or any other polynucleotides of the invention. In one embodiment, the recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a nucleic acid having any of the nucleotide sequences of the SEQ ID NOs: 1 - 35 or a fragment thereof is inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example. Bacterial: pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Eukaryotic: pWLneo, pSV2cat, pOG44, PXTI, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia).

The isolated polynucleotide of the invention may be operably linked to an expression control sequence such as the pMT2 or pED expression vectors disclosed in Kaufman et al., *Nucleic Acids Res.* 19, 4485-4490 (1991), in order to produce the protein recombinantly. Many suitable expression control sequences are known in the art. General methods of expressing recombinant proteins are also known and are exemplified in R. Kaufman, *Methods in Enzymology* 185, 537-566 (1990). As defined herein

"operably linked" means that the isolated polynucleotide of the invention and an expression control sequence are situated within a vector or cell in such a way that the protein is expressed by a host cell which has been transformed (transfected) with the ligated polynucleotide/expression control sequence.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, *e.g.*, the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), a-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an amino terminal identification peptide imparting desired characteristics, *e.g.*, stabilization or simplified purification of expressed recombinant product. Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (Promega Biotech, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed. Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced or derepressed by appropriate means (*e.g.*, temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Polynucleotides of the invention can also be used to induce immune responses. For example, as described in Fan et al., *Nat. Biotech.* 17:870-872 (1999), incorporated herein by reference, nucleic acid sequences encoding a polypeptide may be used to generate antibodies against the encoded polypeptide following topical administration of naked plasmid DNA or following injection, and preferably intramuscular injection of the DNA. The nucleic acid sequences are preferably inserted in a recombinant expression vector and may be in the form of naked DNA.

4.3 HOSTS

The present invention further provides host cells genetically engineered to contain the polynucleotides of the invention. For example, such host cells may contain nucleic acids of the invention introduced into the host cell using known transformation, transfection or infection methods. The present invention still further provides host cells genetically engineered to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell.

Knowledge of nucleic acid sequences allows for modification of cells to permit, or increase, expression of endogenous polypeptide. Cells can be modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in

whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the polypeptide at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the encoding sequences. See, for example, PCT International Publication No. WO94/12650, PCT International Publication No. WO92/20808, and PCT International Publication No. WO91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., ada, dhfr, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the recombinant construct into the host cell can be effected by calcium phosphate transfection, DEAE, dextran mediated transfection, or electroporation (Davis, L. et al., *Basic Methods in Molecular Biology* (1986)). The host cells containing one of the polynucleotides of the invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, Cv-1 cell, COS cells, 293 cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level. Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., in Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell 23:175 (1981). Other cell lines capable of expressing a compatible vector are, for example, the C127, monkey COS cells, Chinese Hamster Ovary (CHO) cells, human kidney 293 cells, human epidermal A431 cells, human Colo205 cells, 3T3 cells, CV-1 cells, other transformed primate cell lines, normal diploid cells, cell strains derived from *in vitro* culture of primary tissue, primary explants, HeLa cells, mouse L cells, BHK, HL-60, U937, HaK or Jurkat cells. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5′ flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements. Recombinant polypeptides and proteins produced in bacterial culture are usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents.

Alternatively, it may be possible to produce the protein in lower eukaryotes such as yeast or insects or in prokaryotes such as bacteria. Potentially suitable yeast strains include *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Kluyveromyces* strains, *Candida*, or any yeast strain capable of expressing heterologous proteins. Potentially suitable bacterial strains include *Escherichia coli, Bacillus subtilis, Salmonella typhimurium*, or any bacterial strain capable of expressing heterologous proteins. If the protein is made in yeast or bacteria, it may be necessary to modify the protein produced therein, for example by phosphorylation or glycosylation of the appropriate sites, in order to obtain the functional protein. Such covalent attachments may be accomplished using known chemical or enzymatic methods.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequence include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, *e.g.,* inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the host cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result

in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

4.4 POLYPEPTIDES OF THE INVENTION

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising: the amino acid sequences set forth as any one of SEQ ID NO: 1-35 or an amino acid sequence encoded by any one of the nucleotide sequences SEQ ID NOs: 1 - 35 or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides preferably with biological or immunological activity that are encoded by: (a) a polynucleotide having any one of the nucleotide sequences set forth in the SEQ ID NOs: 1 – 35 or (b) polynucleotides encoding any one of the amino acid sequences set forth as SEQ ID NO: 1-35 or (c) polynucleotides that hybridize to the complement of the polynucleotides of either (a) or (b) under stringent hybridization conditions. The invention also provides biologically active or immunologically active variants of any of the amino acid sequences set forth as SEQ ID NO: 1-35 or the corresponding full length or mature protein; and "substantial equivalents" thereof (e.g., with at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, typically at least about 95%, more typically at least about 98%, or most typically at least about 99% amino acid identity) that retain biological activity. Polypeptides encoded by allelic variants may have a similar, increased, or decreased activity compared to polypeptides comprising SEQ ID NO: 1-35.

Fragments of the proteins of the present invention which are capable of exhibiting biological activity are also encompassed by the present invention. Fragments of the protein may be in linear form or they may be cyclized using known methods, for

example, as described in H. U. Saragovi, et al., Bio/Technology 10, 773-778 (1992) and in R. S. McDowell, et al., J. Amer. Chem. Soc. 114, 9245-9253 (1992), both of which are incorporated herein by reference. Such fragments may be fused to carrier molecules such as immunoglobulins for many purposes, including increasing the valency of protein binding sites.

The present invention also provides both full-length and mature forms (for example, without a signal sequence or precursor sequence) of the disclosed proteins. The protein coding sequence is identified in the sequence listing by translation of the disclosed nucleotide sequences. The mature form of such protein may be obtained by expression of a full-length polynucleotide in a suitable mammalian cell or other host cell. The sequence of the mature form of the protein is also determinable from the amino acid sequence of the full-length form. Where proteins of the present invention are membrane bound, soluble forms of the proteins are also provided. In such forms, part or all of the regions causing the proteins to be membrane bound are deleted so that the proteins are fully secreted from the cell in which it is expressed.

Protein compositions of the present invention may further comprise an acceptable carrier, such as a hydrophilic, *e.g.*, pharmaceutically acceptable, carrier.

The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ from a nucleic acid fragment of the present invention (*e.g.*, an ORF) by nucleotide sequence but, due to the degeneracy of the genetic code, encode an identical polypeptide sequence. Preferred nucleic acid fragments of the present invention are the ORFs that encode proteins.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. The synthetically-constructed protein sequences, by virtue of sharing primary, secondary or tertiary structural and/or conformational characteristics with proteins may possess biological properties in common therewith, including protein activity. This technique is particularly useful in producing small peptides and fragments

of larger polypeptides. Fragments are useful, for example, in generating antibodies against the native polypeptide. Thus, they may be employed as biologically active or immunological substitutes for natural, purified proteins in screening of therapeutic compounds and in immunological processes for the development of antibodies.

The polypeptides and proteins of the present invention can alternatively be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. One skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

The invention also relates to methods for producing a polypeptide comprising growing a culture of host cells of the invention in a suitable culture medium, and purifying the protein from the cells or the culture in which the cells are grown. For example, the methods of the invention include a process for producing a polypeptide in which a host cell containing a suitable expression vector that includes a polynucleotide of the invention is cultured under conditions that allow expression of the encoded polypeptide. The polypeptide can be recovered from the culture, conveniently from the culture medium, or from a lysate prepared from the host cells and further purified. Preferred embodiments include those in which the protein produced by such process is a full length or mature form of the protein.

In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily follow known methods for isolating polypeptides and proteins in order to obtain one of the isolated polypeptides or proteins of the present invention. These include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immuno-affinity chromatography. See, *e.g.*, Scopes, *Protein Purification: Principles and Practice*, Springer-Verlag (1994); Sambrook, et al., in Molecular Cloning: *A Laboratory Manual*; Ausubel et al., *Current Protocols in*

Molecular Biology. Polypeptide fragments that retain biological/immunological activity include fragments comprising greater than about 100 amino acids, or greater than about 200 amino acids, and fragments that encode specific protein domains.

The purified polypeptides can be used in *in vitro* binding assays which are well known in the art to identify molecules which bind to the polypeptides. These molecules include but are not limited to, for e.g., small molecules, molecules from combinatorial libraries, antibodies or other proteins. The molecules identified in the binding assay are then tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

In addition, the peptides of the invention or molecules capable of binding to the peptides may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for SEQ ID NO: 1-35.

The protein of the invention may also be expressed as a product of transgenic animals, *e.g.,* as a component of the milk of transgenic cows, goats, pigs, or sheep which are characterized by somatic or germ cells containing a nucleotide sequence encoding the protein.

The proteins provided herein also include proteins characterized by amino acid sequences similar to those of purified proteins but into which modification are naturally provided or deliberately engineered. For example, modifications, in the peptide or DNA sequence, can be made by those skilled in the art using known techniques. Modifications of interest in the protein sequences may include the alteration, substitution, replacement, insertion or deletion of a selected amino acid residue in the coding sequence. For example, one or more of the cysteine residues may be deleted or replaced with another amino acid to alter the conformation of the molecule. Techniques for such alteration, substitution, replacement, insertion or deletion are well known to those skilled in the art (see, *e.g.,* U.S. Pat. No. 4,518,584). Preferably, such alteration, substitution, replacement, insertion or deletion retains the desired activity of the protein. Regions of the protein that are important for the protein function can be determined by various methods known in

the art including the alanine-scanning method which involved systematic substitution of single or strings of amino acids with alanine, followed by testing the resulting alanine-containing variant for biological activity. This type of analysis determines the importance of the substituted amino acid(s) in biological activity. Regions of the protein that are important for protein function may be determined by the eMATRIX program.

Other fragments and derivatives of the sequences of proteins which would be expected to retain protein activity in whole or in part and are useful for screening or other immunological methodologies may also be easily made by those skilled in the art given the disclosures herein. Such modifications are encompassed by the present invention.

The protein may also be produced by operably linking the isolated polynucleotide of the invention to suitable control sequences in one or more insect expression vectors, and employing an insect expression system. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *e.g.*, Invitrogen, San Diego, Calif., U.S.A. (the MaxBat™ kit), and such methods are well known in the art, as described in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987), incorporated herein by reference. As used herein, an insect cell capable of expressing a polynucleotide of the present invention is "transformed."

The protein of the invention may be prepared by culturing transformed host cells under culture conditions suitable to express the recombinant protein. The resulting expressed protein may then be purified from such culture (*i.e.*, from culture medium or cell extracts) using known purification processes, such as gel filtration and ion exchange chromatography. The purification of the protein may also include an affinity column containing agents which will bind to the protein; one or more column steps over such affinity resins as concanavalin A-agarose, heparin-toyopearl™ or Cibacrom blue 3GA Sepharose™; one or more steps involving hydrophobic interaction chromatography using such resins as phenyl ether, butyl ether, or propyl ether; or immunoaffinity chromatography.

Alternatively, the protein of the invention may also be expressed in a form which will facilitate purification. For example, it may be expressed as a fusion protein, such as those of maltose binding protein (MBP), glutathione-S-transferase (GST) or thioredoxin

(TRX), or as a His tag. Kits for expression and purification of such fusion proteins are commercially available from New England BioLab (Beverly, Mass.), Pharmacia (Piscataway, N.J.) and Invitrogen, respectively. The protein can also be tagged with an epitope and subsequently purified by using a specific antibody directed to such epitope. One such epitope ("FLAG®") is commercially available from Kodak (New Haven, Conn.).

Finally, one or more reverse-phase high performance liquid chromatography (RP-HPLC) steps employing hydrophobic RP-HPLC media, *e.g.*, silica gel having pendant methyl or other aliphatic groups, can be employed to further purify the protein. Some or all of the foregoing purification steps, in various combinations, can also be employed to provide a substantially homogeneous isolated recombinant protein. The protein thus purified is substantially free of other mammalian proteins and is defined in accordance with the present invention as an "isolated protein."

The polypeptides of the invention include analogs (variants). This embraces fragments, as well as peptides in which one or more amino acids has been deleted, inserted, or substituted. Also, analogs of the polypeptides of the invention embrace fusions of the polypeptides or modifications of the polypeptides of the invention, wherein the polypeptide or analog is fused to another moiety or moieties, e.g., targeting moiety or another therapeutic agent. Such analogs may exhibit improved properties such as activity and/or stability. Examples of moieties which may be fused to the polypeptide or an analog include, for example, targeting moieties which provide for the delivery of polypeptide to pancreatic cells, e.g., antibodies to pancreatic cells, antibodies to immune cells such as T-cells, monocytes, dendritic cells, granulocytes, etc., as well as receptor and ligands expressed on pancreatic or immune cells. Other moieties which may be fused to the polypeptide include therapeutic agents which are used for treatment, for example, immunosuppressive drugs such as cyclosporin, SK506, azathioprine, CD3 antibodies and steroids. Also, polypeptides may be fused to immune modulators, and other cytokines such as alpha or beta interferon.

4.4.1 DETERMINING POLYPEPTIDE AND POLYNUCLEOTIDE IDENTITY AND SIMILARITY

Preferred identity and/or similarity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in computer programs including, but are not limited to, the GCG program package, including GAP (Devereux, J., et al., Nucleic Acids Research 12(1):387 (1984); Genetics Computer Group, University of Wisconsin, Madison, WI), BLASTP, BLASTN, BLASTX, FASTA (Altschul, S.F. et al., J. Molec. Biol. 215:403-410 (1990), PSI-BLAST (Altschul S.F. et al., Nucleic Acids Res. vol. 25, pp. 3389-3402, herein incorporated by reference), eMatrix software (Wu et al., J. Comp. Biol., Vol. 6, pp. 219-235 (1999), herein incorporated by reference), eMotif software (Nevill-Manning et al, ISMB-97, Vol. 4, pp. 202-209, herein incorporated by reference), pFam software (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1), pp. 320-322 (1998), herein incorporated by reference) and the Kyte-Doolittle hydrophobocity prediction algorithm (J. Mol Biol, 157, pp. 105-31 (1982), incorporated herein by reference). The BLAST programs are publicly available from the National Center for Biotechnology Information (NCBI) and other sources (BLAST Manual, Altschul, S., et al. NCB NLM NIH Bethesda, MD 20894; Altschul, S., et al., J. Mol. Biol. 215:403-410 (1990).

4.5 GENE THERAPY

Mutations in the polynucleotides of the invention gene may result in loss of normal function of the encoded protein. The invention thus provides gene therapy to restore normal activity of the polypeptides of the invention; or to treat disease states involving polypeptides of the invention. Delivery of a functional gene encoding polypeptides of the invention to appropriate cells is effected *ex vivo*, *in situ*, or *in vivo* by use of vectors, and more particularly viral vectors (e.g., adenovirus, adeno-associated virus, or a retrovirus), or *ex vivo* by use of physical DNA transfer methods (e.g., liposomes or chemical treatments). See, for example, Anderson, Nature, supplement to vol. 392, no. 6679, pp.25-20 (1998). For additional reviews of gene therapy technology see Friedmann, Science, 244: 1275-1281 (1989); Verma, Scientific American: 68-84 (1990); and Miller, Nature, 357: 455-460 (1992). Introduction of any one of the nucleotides of the present invention or a gene encoding the polypeptides of the present invention can also be accomplished with extrachromosomal substrates (transient

expression) or artificial chromosomes (stable expression). Cells may also be cultured *ex vivo* in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced *in vivo* for therapeutic purposes. Alternatively, it is contemplated that in other human disease states, preventing the expression of or inhibiting the activity of polypeptides of the invention will be useful in treating the disease states. It is contemplated that antisense therapy or gene therapy could be applied to negatively regulate the expression of polypeptides of the invention.

Other methods inhibiting expression of a protein include the introduction of antisense molecules to the nucleic acids of the present invention, their complements, or their translated RNA sequences, by methods known in the art. Further, the polypeptides of the present invention can be inhibited by using targeted deletion methods, or the insertion of a negative regulatory element such as a silencer, which is tissue specific.

The present invention still further provides cells genetically engineered *in vivo* to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell. These methods can be used to increase or decrease the expression of the polynucleotides of the present invention.

Knowledge of DNA sequences provided by the invention allows for modification of cells to permit, increase, or decrease, expression of endogenous polypeptide. Cells can be modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the protein at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the desired protein encoding sequences. See, for example, PCT International Publication No. WO 94/12650, PCT International Publication No. WO 92/20808, and PCT International Publication No. WO 91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., ada, dhfr, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the desired

protein coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequences include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, *e.g.,* inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting

sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

4.6 TRANSGENIC ANIMALS

In preferred methods to determine biological functions of the polypeptides of the invention in vivo, one or more genes provided by the invention are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of a promoter of the polynucleotides of the invention is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or even replacing the homologous promoter to provide for increased

protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

The polynucleotides of the present invention also make possible the development, through, e.g., homologous recombination or knock out strategies, of animals that fail to express polypeptides of the invention or that express a variant polypeptide. Such animals are useful as models for studying the *in vivo* activities of polypeptide as well as for studying modulators of the polypeptides of the invention.

In preferred methods to determine biological functions of the polypeptides of the invention *in vivo*, one or more genes provided by the invention are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of the polynucleotides of the invention promoter is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or even replacing the homologous promoter to provide for increased protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

4.7 USES AND BIOLOGICAL ACTIVITY

The polynucleotides and proteins of the present invention are expected to exhibit one or more of the uses or biological activities (including those associated with assays cited herein) identified herein. Uses or activities described for proteins of the present invention may be provided by administration or use of such proteins or of polynucleotides encoding such proteins (such as, for example, in gene therapies or vectors suitable for introduction of DNA). The mechanism underlying the particular condition or pathology will dictate whether the polypeptides of the invention, the polynucleotides of the invention or modulators (activators or inhibitors) thereof would be beneficial to the subject in need of treatment. Thus, "therapeutic compositions of the invention" include compositions comprising isolated polynucleotides (including recombinant DNA molecules, cloned genes and degenerate variants thereof) or polypeptides of the invention (including full length protein, mature protein and truncations or domains thereof), or compounds and other substances that modulate the overall activity of the target gene products, either at the level of target gene/protein expression or target protein activity. Such modulators include polypeptides, analogs, (variants), including fragments and fusion proteins, antibodies and other binding proteins; chemical compounds that directly or indirectly activate or inhibit the polypeptides of the invention (identified, e.g., via drug screening assays as described herein); antisense polynucleotides and polynucleotides suitable for triple helix formation; and in particular antibodies or other binding partners that specifically recognize one or more epitopes of the polypeptides of the invention.

The polypeptides of the present invention may likewise be involved in cellular activation or in one of the other physiological pathways described herein.

4.7.1 RESEARCH USES AND UTILITIES

The polynucleotides provided by the present invention can be used by the research community for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on gels; as chromosome markers or tags (when

labeled) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; as a probe to "subtract-out" known sequences in the process of discovering other novel polynucleotides; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination of expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris et al., Cell 75:791-803 (1993)) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

The polypeptides provided by the present invention can similarly be used in assays to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding polypeptide is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Proteins involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology:

Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

4.7.2 NUTRITIONAL USES

Polynucleotides and polypeptides of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or amino acid supplement, use as a carbon source, use as a nitrogen source and use as a source of carbohydrate. In such cases the polypeptide or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the polypeptide or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

4.7.3 CYTOKINE AND CELL PROLIFERATION/DIFFERENTIATION ACTIVITY

A polypeptide of the present invention may exhibit activity relating to cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor-dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of therapeutic compositions of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+(preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e, CMK, HUVEC, and Caco. Therapeutic compositions of the invention can be used in the following:

Assays for T-cell or thymocyte proliferation include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function 3.1-3.19;

Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Bertagnolli et al., J. Immunol. 145:1706-1712, 1990; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Bertagnolli, et al., I. Immunol. 149:3778-3783, 1992; Bowman et al., I. Immunol. 152:1756-1761, 1994.

Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described in: Polyclonal T cell stimulation, Kruisbeek, A. M. and Shevach, E. M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.12.1-3.12.14, John Wiley and Sons, Toronto. 1994; and Measurement of mouse and human interleukin- , Schreiber, R. D. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described in: Measurement of Human and Murine Interleukin 2 and Interleukin 4, Bottomly, K., Davis, L. S. and Lipsky, P. E. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries et al., J. Exp. Med. 173:1205-1211, 1991; Moreau et al., Nature 336:690-692, 1988; Greenberger et al., Proc. Natl. Acad. Sci. U.S.A. 80:2931-2938, 1983; Measurement of mouse and human interleukin 6--Nordan, R. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto. 1991; Smith et al., Proc. Natl. Aced. Sci. U.S.A. 83:1857-1861, 1986; Measurement of human Interleukin 11--Bennett, F., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto. 1991; Measurement of mouse and human Interleukin 9--Ciarletta, A., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto. 1991.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and cytokine production) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function; Chapter

6, Cytokines and their cellular receptors; Chapter 7, Immunologic studies in Humans); Weinberger et al., Proc. Natl. Acad. Sci. USA 77:6091-6095, 1980; Weinberger et al., Eur. J. Immun. 11:405-411, 1981; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988.

4.7.4 STEM CELL GROWTH FACTOR ACTIVITY

A polypeptide of the present invention may exhibit stem cell growth factor activity and be involved in the proliferation, differentiation and survival of pluripotent and totipotent stem cells including primordial germ cells, embryonic stem cells, hematopoietic stem cells and/or germ line stem cells. Administration of the polypeptide of the invention to stem cells *in vivo* or *ex vivo* is expected to maintain and expand cell populations in a totipotential or pluripotential state which would be useful for re-engineering damaged or diseased tissues, transplantation, manufacture of bio-pharmaceuticals and the development of bio-sensors. The ability to produce large quantities of human cells has important working applications for the production of human proteins which currently must be obtained from non-human sources or donors, implantation of cells to treat diseases such as Parkinson's, Alzheimer's and other neurodegenerative diseases; tissues for grafting such as bone marrow, skin, cartilage, tendons, bone, muscle (including cardiac muscle), blood vessels, cornea, neural cells, gastrointestinal cells and others; and organs for transplantation such as kidney, liver, pancreas (including islet cells), heart and lung.

It is contemplated that multiple different exogenous growth factors and/or cytokines may be administered in combination with the polypeptide of the invention to achieve the desired effect, including any of the growth factors listed herein, other stem cell maintenance factors, and specifically including stem cell factor (SCF), leukemia inhibitory factor (LIF), Flt-3 ligand (Flt-3L), any of the interleukins, recombinant soluble IL-6 receptor fused to IL-6, macrophage inflammatory protein 1-alpha (MIP-1-alpha), G-CSF, GM-CSF, thrombopoietin (TPO), platelet factor 4 (PF-4), platelet-derived growth factor (PDGF), neural growth factors and basic fibroblast growth factor (bFGF).

Since totipotent stem cells can give rise to virtually any mature cell type, expansion of these cells in culture will facilitate the production of large quantities of

mature cells. Techniques for culturing stem cells are known in the art and administration of polypeptides of the invention, optionally with other growth factors and/or cytokines, is expected to enhance the survival and proliferation of the stem cell populations. This can be accomplished by direct administration of the polypeptide of the invention to the culture medium. Alternatively, stroma cells transfected with a polynucleotide that encodes for the polypeptide of the invention can be used as a feeder layer for the stem cell populations in culture or in vivo. Stromal support cells for feeder layers may include embryonic bone marrow fibroblasts, bone marrow stromal cells, fetal liver cells, or cultured embryonic fibroblasts (see U.S. Patent No. 5,690,926).

Stem cells themselves can be transfected with a polynucleotide of the invention to induce autocrine expression of the polypeptide of the invention. This will allow for generation of undifferentiated totipotential/pluripotential stem cell lines that are useful as is or that can then be differentiated into the desired mature cell types. These stable cell lines can also serve as a source of undifferentiated totipotential/pluripotential mRNA to create cDNA libraries and templates for polymerase chain reaction experiments. These studies would allow for the isolation and identification of differentially expressed genes in stem cell populations that regulate stem cell proliferation and/or maintenance.

Expansion and maintenance of totipotent stem cell populations will be useful in the treatment of many pathological conditions. For example, polypeptides of the present invention may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease, accidental damage or genetic disorders. The polypeptide of the invention may be useful for inducing the proliferation of neural cells and for the regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders which involve degeneration, death or trauma to neural cells or nerve tissue. In addition, the expanded stem cell populations can also be genetically altered for gene therapy purposes and to decrease host rejection of replacement tissues after grafting or implantation.

Expression of the polypeptide of the invention and its effect on stem cells can also be manipulated to achieve controlled differentiation of the stem cells into more differentiated cell types. A broadly applicable method of obtaining pure populations of a

specific differentiated cell type from undifferentiated stem cell populations involves the use of a cell-type specific promoter driving a selectable marker. The selectable marker allows only cells of the desired type to survive. For example, stem cells can be induced to differentiate into cardiomyocytes (Wobus et al., Differentiation, 48: 173-182, (1991); Klug et al., J. Clin. Invest., 98(1): 216-224, (1998)) or skeletal muscle cells (Browder, L. W. In: *Principles of Tissue Engineering eds.* Lanza et al., Academic Press (1997)). Alternatively, directed differentiation of stem cells can be accomplished by culturing the stem cells in the presence of a differentiation factor such as retinoic acid and an antagonist of the polypeptide of the invention which would inhibit the effects of endogenous stem cell factor activity and allow differentiation to proceed.

In vitro cultures of stem cells can be used to determine if the polypeptide of the invention exhibits stem cell growth factor activity. Stem cells are isolated from any one of various cell sources (including hematopoietic stem cells and embryonic stem cells) and cultured on a feeder layer, as described by Thompson et al. Proc. Natl. Acad. Sci, U.S.A., 92: 7844-7848 (1995), in the presence of the polypeptide of the invention alone or in combination with other growth factors or cytokines. The ability of the polypeptide of the invention to induce stem cells proliferation is determined by colony formation on semi-solid support e.g. as described by Bernstein et al., Blood, 77: 2316-2321 (1991).

4.7.5 HEMATOPOIESIS REGULATING ACTIVITY

A polypeptide of the present invention may be involved in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell disorders. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation

of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either *in-vivo* or *ex-vivo* (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy.

Therapeutic compositions of the invention can be used in the following:

Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson et al. Cellular Biology 15:141-151, 1995; Keller et al., Molecular and Cellular Biology 13:473-486, 1993; McClanahan et al., Blood 81:2903-2915, 1993.

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, M. G. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. 1994; Hirayama et al., Proc. Natl. Acad. Sci. USA 89:5907-5911, 1992; Primitive hematopoietic colony forming cells with high proliferative potential, McNiece, I. K. and Briddell, R. A. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. 1994; Neben et al., Experimental Hematology 22:353-359, 1994; Cobblestone area forming cell assay, Ploemacher, R. E. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. 1994; Long term bone marrow cultures in the presence of stromal cells, Spooncer, E., Dexter, M. and Allen, T. In Culture of Hematopoietic Cells. R. I.

Freshney, et al. eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Long term culture initiating cell assay, Sutherland, H. J. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 139-162, Wiley-Liss, Inc., New York, N.Y. 1994.

4.7.6 TISSUE GROWTH ACTIVITY

A polypeptide of the present invention also may be involved in bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as in wound healing and tissue repair and replacement, and in healing of burns, incisions and ulcers.

A polypeptide of the present invention which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Compositions of a polypeptide, antibody, binding partner, or other modulator of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A polypeptide of this invention may also be involved in attracting bone-forming cells, stimulating growth of bone-forming cells, or inducing differentiation of progenitors of bone-forming cells. Treatment of osteoporosis, osteoarthritis, bone degenerative disorders, or periodontal disease, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes may also be possible using the composition of the invention.

Another category of tissue regeneration activity that may involve the polypeptide of the present invention is tendon/ligament formation. Induction of tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue.

De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex vivo* for return *in vivo* to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The compositions of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a composition may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a composition of the invention.

Compositions of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

Compositions of the present invention may also be involved in the generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising

such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring may allow normal tissue to regenerate. A polypeptide of the present invention may also exhibit angiogenic activity.

A composition of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A composition of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

Therapeutic compositions of the invention can be used in the following:

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in: Winter, Epidermal Wound Healing, pps. 71-112 (Maibach, H. I. and Rovee, D. T., eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dermatol 71:382-84 (1978).

4.7.7 IMMUNE STIMULATING OR SUPPRESSING ACTIVITY

A polypeptide of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A polynucleotide of the invention can encode a polypeptide exhibiting such activities. A protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency (SCID)), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases causes by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpes

viruses, mycobacteria, Leishmania spp., malaria spp. and various fungal infections such as candidiasis. Of course, in this regard, proteins of the present invention may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein (or antagonists thereof, including antibodies) of the present invention may also to be useful in the treatment of allergic reactions and conditions (*e.g.*, anaphylaxis, serum sickness, drug reactions, food allergies, insect venom allergies, mastocytosis, allergic rhinitis, hypersensitivity pneumonitis, urticaria, angioedema, eczema, atopic dermatitis, allergic contact dermatitis, erythema multiforme, Stevens-Johnson syndrome, allergic conjunctivitis, atopic keratoconjunctivitis, venereal keratoconjunctivitis, giant papillary conjunctivitis and contact allergies), such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein (or antagonists thereof) of the present invention. The therapeutic effects of the polypeptides or antagonists thereof on allergic reactions can be evaluated by in vivo animals models such as the cumulative contact enhancement test (Lastbom et al., Toxicology 125: 59-66, 1998), skin prick test (Hoffmann et al., Allergy 54: 446-54, 1999), guinea pig skin sensitization test (Vohr et al., Arch. Toxocol. 73: 501-9), and murine local lymph node assay (Kimber et al., J. Toxicol. Environ. Health 53: 563-79).

Using the proteins of the invention it may also be possible to modulate immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response. The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent.

Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as, for example, B7)), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a therapeutic composition of the invention may prevent cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, a lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular therapeutic compositions in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins in vivo as described in Lenschow et al., Science 257:789-792 (1992) and Turka et al., Proc. Natl. Acad. Sci USA, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of therapeutic compositions of the invention on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block stimulation of T cells can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythmatosis in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (e.g., a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response may be useful in cases of viral infection, including systemic viral diseases such as influenza, the common cold, and encephalitis.

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells in vitro with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the in vitro activated T cells into the patient. Another method of enhancing anti-viral immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected

cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells in vivo.

A polypeptide of the present invention may provide the necessary stimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient mounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I alpha chain protein and β_2 microglobulin protein or an MHC class II alpha chain protein and an MHC class II beta chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 1981; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J. Immunol. 135:1564-1572, 1985; Takai et al., I. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bowman et al., J. Virology 61:1992-1998; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Brown et al., J. Immunol. 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, J. Immunol. 144:3028-3033, 1990; and Assays for B cell function: In vitro antibody production, Mond, J. J. and Brunswick, M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto. 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bertagnolli et al., J. Immunol. 149:3778-3783, 1992.

Dendritic cell-dependent assays (which will identify, among others, proteins expressed by dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery et al., J. Immunol. 134:536-544, 1995; Inaba et al., Journal of Experimental Medicine 173:549-559, 1991; Macatonia et al., Journal of Immunology 154:5071-5079, 1995; Porgador et al., Journal of Experimental Medicine 182:255-260, 1995; Nair et al., Journal of Virology 67:4062-4069, 1993; Huang et al., Science 264:961-965, 1994; Macatonia et al., Journal of Experimental Medicine 169:1255-1264, 1989; Bhardwaj et al., Journal of Clinical Investigation 94:797-807, 1994; and Inaba et al., Journal of Experimental Medicine 172:631-640, 1990.

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz et al., Cytometry 13:795-808, 1992; Gorczyca et al., Leukemia 7:659-670, 1993; Gorczyca et al., Cancer Research 53:1945-1951, 1993; Itoh et al., Cell 66:233-243, 1991; Zacharchuk, Journal of Immunology 145:4037-4045, 1990; Zamai et al., Cytometry 14:891-897, 1993; Gorczyca et al., International Journal of Oncology 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica et al., Blood 84:111-117, 1994; Fine et al., Cellular Immunology 155:111-122, 1994; Galy et al., Blood 85:2770-2778, 1995; Toki et al., Proc. Nat. Acad Sci. USA 88:7548-7551, 1991.

4.7.8 ACTIVIN/INHIBIN ACTIVITY

A polypeptide of the present invention may also exhibit activin- or inhibin-related activities. A polynucleotide of the invention may encode a polypeptide exhibiting such characteristics. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins and are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a polypeptide of the present invention, alone or in heterodimers with a member of the inhibin family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the polypeptide of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, U.S. Pat. No. 4,798,885. A polypeptide of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as, but not limited to, cows, sheep and pigs.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods.

Assays for activin/inhibin activity include, without limitation, those described in: Vale et al., Endocrinology 91:562-572, 1972; Ling et al., Nature 321:779-782, 1986; Vale et al., Nature 321:776-779, 1986; Mason et al., Nature 318:659-663, 1985; Forage et al., Proc. Natl. Acad. Sci. USA 83:3091-3095, 1986.

4.7.9 CHEMOTACTIC/CHEMOKINETIC ACTIVITY

A polypeptide of the present invention may be involved in chemotactic or chemokinetic activity for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Chemotactic and chemokinetic receptor activation can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic compositions (e.g. proteins, antibodies, binding partners, or modulators of the invention) provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

Therapeutic compositions of the invention can be used in the following:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Marguiles, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub et al. J. Clin. Invest. 95:1370-1376, 1995; Lind et al. APMIS 103:140-146, 1995; Muller et al Eur. J. Immunol. 25:1744-1748; Gruber et al. J. of Immunol. 152:5860-5867, 1994; Johnston et al. J. of Immunol. 153:1762-1768, 1994.

4.7.10 HEMOSTATIC AND THROMBOLYTIC ACTIVITY

A polypeptide of the invention may also be involved in hemostatis or thrombolysis or thrombosis. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Compositions may be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A composition of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke).

Therapeutic compositions of the invention can be used in the following:

Assay for hemostatic and thrombolytic activity include, without limitation, those described in: Linet et al., J. Clin. Pharmacol. 26:131-140, 1986; Burdick et al., Thrombosis Res. 45:413-419, 1987; Humphrey et al., Fibrinolysis 5:71-79 (1991); Schaub, Prostaglandins 35:467-474, 1988.

4.7.11 CANCER DIAGNOSIS AND THERAPY

Polypeptides of the invention may be involved in cancer cell generation, proliferation or metastasis. Detection of the presence or amount of polynucleotides or polypeptides of the invention may be useful for the diagnosis and/or prognosis of one or more types of cancer. For example, the presence or increased expression of a polynucleotide/polypeptide of the invention may indicate a hereditary risk of cancer, a precancerous condition, or an ongoing malignancy. Conversely, a defect in the gene or absence of the polypeptide may be associated with a cancer condition. Identification of single nucleotide polymorphisms associated with cancer or a predisposition to cancer may also be useful for diagnosis or prognosis.

Cancer treatments promote tumor regression by inhibiting tumor cell proliferation, inhibiting angiogenesis (growth of new blood vessels that is necessary to support tumor growth) and/or prohibiting metastasis by reducing tumor cell motility or invasiveness. Therapeutic compositions of the invention may be effective in adult and pediatric oncology including in solid phase tumors/malignancies, locally advanced tumors, human soft tissue sarcomas, metastatic cancer, including lymphatic metastases,

blood cell malignancies including multiple myeloma, acute and chronic leukemias, and lymphomas, head and neck cancers including mouth cancer, larynx cancer and thyroid cancer, lung cancers including small cell carcinoma and non-small cell cancers, breast cancers including small cell carcinoma and ductal carcinoma, gastrointestinal cancers including esophageal cancer, stomach cancer, colon cancer, colorectal cancer and polyps associated with colorectal neoplasia, pancreatic cancers, liver cancer, urologic cancers including bladder cancer and prostate cancer, malignancies of the female genital tract including ovarian carcinoma, uterine (including endometrial) cancers, and solid tumor in the ovarian follicle, kidney cancers including renal cell carcinoma, brain cancers including intrinsic brain tumors, neuroblastoma, astrocytic brain tumors, gliomas, metastatic tumor cell invasion in the central nervous system, bone cancers including osteomas, skin cancers including malignant melanoma, tumor progression of human skin keratinocytes, squamous cell carcinoma, basal cell carcinoma, hemangiopericytoma and Karposi's sarcoma.

Polypeptides, polynucleotides, or modulators of polypeptides of the invention (including inhibitors and stimulators of the biological activity of the polypeptide of the invention) may be administered to treat cancer. Therapeutic compositions can be administered in therapeutically effective dosages alone or in combination with adjuvant cancer therapy such as surgery, chemotherapy, radiotherapy, thermotherapy, and laser therapy, and may provide a beneficial effect, e.g. reducing tumor size, slowing rate of tumor growth, inhibiting metastasis, or otherwise improving overall clinical condition, without necessarily eradicating the cancer.

The composition can also be administered in therapeutically effective amounts as a portion of an anti-cancer cocktail. An anti-cancer cocktail is a mixture of the polypeptide or modulator of the invention with one or more anti-cancer drugs in addition to a pharmaceutically acceptable carrier for delivery. The use of anti-cancer cocktails as a cancer treatment is routine. Anti-cancer drugs that are well known in the art and can be used as a treatment in combination with the polypeptide or modulator of the invention include: Actinomycin D, Aminoglutethimide, Asparaginase, Bleomycin, Busulfan, Carboplatin, Carmustine, Chlorambucil, Cisplatin (cis-DDP), Cyclophosphamide, Cytarabine HCl (Cytosine arabinoside), Dacarbazine, Dactinomycin, Daunorubicin HCl,

Doxorubicin HCl, Estramustine phosphate sodium, Etoposide (V16-213), Floxuridine, 5-Fluorouracil (5-Fu), Flutamide, Hydroxyurea (hydroxycarbamide), Ifosfamide, Interferon Alpha-2a, Interferon Alpha-2b, Leuprolide acetate (LHRH-releasing factor analog), Lomustine, Mechlorethamine HCl (nitrogen mustard), Melphalan, Mercaptopurine, Mesna, Methotrexate (MTX), Mitomycin, Mitoxantrone HCl, Octreotide, Plicamycin, Procarbazine HCl, Streptozocin, Tamoxifen citrate, Thioguanine, Thiotepa, Vinblastine sulfate, Vincristine sulfate, Amsacrine, Azacitidine, Hexamethylmelamine, Interleukin-2, Mitoguazone, Pentostatin, Semustine, Teniposide, and Vindesine sulfate.

In addition, therapeutic compositions of the invention may be used for prophylactic treatment of cancer. There are hereditary conditions and/or environmental situations (e.g. exposure to carcinogens) known in the art that predispose an individual to developing cancers. Under these circumstances, it may be beneficial to treat these individuals with therapeutically effective doses of the polypeptide of the invention to reduce the risk of developing cancers.

In vitro models can be used to determine the effective doses of the polypeptide of the invention as a potential cancer treatment. These *in vitro* models include proliferation assays of cultured tumor cells, growth of cultured tumor cells in soft agar (see Freshney, (1987) Culture of Animal Cells: A Manual of Basic Technique, Wily-Liss, New York, NY Ch 18 and Ch 21), tumor systems in nude mice as described in Giovanella et al., J. Natl. Can. Inst., 52: 921-30 (1974), mobility and invasive potential of tumor cells in Boyden Chamber assays as described in Pilkington et al., Anticancer Res., 17: 4107-9 (1997), and angiogenesis assays such as induction of vascularization of the chick chorioallantoic membrane or induction of vascular endothelial cell migration as described in Ribatta et al., Intl. J. Dev. Biol., 40: 1189-97 (1999) and Li et al., Clin. Exp. Metastasis, 17:423-9 (1999), respectively. Suitable tumor cells lines are available, e.g. from American Type Tissue Culture Collection catalogs.

4.7.12 RECEPTOR/LIGAND ACTIVITY

A polypeptide of the present invention may also demonstrate activity as receptor, receptor ligand or inhibitor or agonist of receptor/ligand interactions. A polynucleotide of the invention can encode a polypeptide exhibiting such characteristics. Examples of

such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses. Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley- Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions 7.28.1- 7.28.22), Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein et al., J. Exp. Med. 169:149-160 1989; Stoltenborg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995.

By way of example, the polypeptides of the invention may be used as a receptor for a ligand(s) thereby transmitting the biological activity of that ligand(s). Ligands may be identified through binding assays, affinity chromatography, dihybrid screening assays, BIAcore assays, gel overlay assays, or other methods known in the art.

Studies characterizing drugs or proteins as agonist or antagonist or partial agonists or a partial antagonist require the use of other proteins as competing ligands. The polypeptides of the present invention or ligand(s) thereof may be labeled by being coupled to radioisotopes, colorimetric molecules or a toxin molecules by conventional methods. ("Guide to Protein Purification" Murray P. Deutscher (ed) Methods in Enzymology Vol. 182 (1990) Academic Press, Inc. San Diego). Examples of radioisotopes include, but are not limited to, tritium and carbon-14 . Examples of

colorimetric molecules include, but are not limited to, fluorescent molecules such as fluorescamine, or rhodamine or other colorimetric molecules. Examples of toxins include, but are not limited, to ricin.

4.7.13 DRUG SCREENING

This invention is particularly useful for screening chemical compounds by using the novel polypeptides or binding fragments thereof in any of a variety of drug screening techniques. The polypeptides or fragments employed in such a test may either be free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant nucleic acids expressing the polypeptide or a fragment thereof. Drugs are screened against such transformed cells in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may measure, for example, the formation of complexes between polypeptides of the invention or fragments and the agent being tested or examine the diminution in complex formation between the novel polypeptides and an appropriate cell line, which are well known in the art.

Sources for test compounds that may be screened for ability to bind to or modulate (i.e., increase or decrease) the activity of polypeptides of the invention include (1) inorganic and organic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of either random or mimetic peptides, oligonucleotides or organic molecules.

Chemical libraries may be readily synthesized or purchased from a number of commercial sources, and may include structural analogs of known compounds or compounds that are identified as "hits" or "leads" via natural product screening.

The sources of natural product libraries are microorganisms (including bacteria and fungi), animals, plants or other vegetation, or marine organisms, and libraries of mixtures for screening may be created by: (1) fermentation and extraction of broths from soil, plant or marine microorganisms or (2) extraction of the organisms themselves. Natural product libraries include polyketides, non-ribosomal peptides, and (non-naturally occurring) variants thereof. For a review, see *Science 282*:63-68 (1998).

Combinatorial libraries are composed of large numbers of peptides, oligonucleotides or organic compounds and can be readily prepared by traditional automated synthesis methods, PCR, cloning or proprietary synthetic methods. Of particular interest are peptide and oligonucleotide combinatorial libraries. Still other libraries of interest include peptide, protein, peptidomimetic, multiparallel synthetic collection, recombinatorial, and polypeptide libraries. For a review of combinatorial chemistry and libraries created therefrom, see Myers, *Curr. Opin. Biotechnol.* 8:701-707 (1997). For reviews and examples of peptidomimetic libraries, see Al-Obeidi et al., *Mol. Biotechnol*, 9(3):205-23 (1998); Hruby et al., *Curr Opin Chem Biol*, 1(1):114-19 (1997); Dorner et al., *Bioorg Med Chem*, 4(5):709-15 (1996) (alkylated dipeptides).

Identification of modulators through use of the various libraries described herein permits modification of the candidate "hit" (or "lead") to optimize the capacity of the "hit" to bind a polypeptide of the invention. The molecules identified in the binding assay are then tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

The binding molecules thus identified may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells such as radioisotopes. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for a polypeptide of the invention. Alternatively, the binding molecules may be complexed with imaging agents for targeting and imaging purposes.

4.7.14 ASSAY FOR RECEPTOR ACTIVITY

The invention also provides methods to detect specific binding of a polypeptide e.g. a ligand or a receptor. The art provides numerous assays particularly useful for identifying previously unknown binding partners for receptor polypeptides of the invention. For example, expression cloning using mammalian or bacterial cells, or dihybrid screening assays can be used to identify polynucleotides encoding binding partners. As another example, affinity chromatography with the appropriate immobilized polypeptide of the invention can be used to isolate polypeptides that recognize and bind

polypeptides of the invention. There are a number of different libraries used for the identification of compounds, and in particular small molecules, that modulate (*i.e.*, increase or decrease) biological activity of a polypeptide of the invention. Ligands for receptor polypeptides of the invention can also be identified by adding exogenous ligands, or cocktails of ligands to two cells populations that are genetically identical except for the expression of the receptor of the invention: one cell population expresses the receptor of the invention whereas the other does not. The response of the two cell populations to the addition of ligands(s) are then compared. Alternatively, an expression library can be co-expressed with the polypeptide of the invention in cells and assayed for an autocrine response to identify potential ligand(s). As still another example, BIAcore assays, gel overlay assays, or other methods known in the art can be used to identify binding partner polypeptides, including, (1) organic and inorganic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of random peptides, oligonucleotides or organic molecules.

The role of downstream intracellular signaling molecules in the signaling cascade of the polypeptide of the invention can be determined. For example, a chimeric protein in which the cytoplasmic domain of the polypeptide of the invention is fused to the extracellular portion of a protein, whose ligand has been identified, is produced in a host cell. The cell is then incubated with the ligand specific for the extracellular portion of the chimeric protein, thereby activating the chimeric receptor. Known downstream proteins involved in intracellular signaling can then be assayed for expected modifications i.e. phosphorylation. Other methods known to those in the art can also be used to identify signaling molecules involved in receptor activity.

4.7.15 ANTI-INFLAMMATORY ACTIVITY

Compositions of the present invention may also exhibit anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or

promote an inflammatory response. Compositions with such activities can be used to treat inflammatory conditions including chronic or acute conditions), including without limitation intimation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Compositions of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material. Compositions of this invention may be utilized to prevent or treat conditions such as, but not limited to, sepsis, acute pancreatitis, endotoxin shock, cytokine induced shock, rheumatoid arthritis, chronic inflammatory arthritis, pancreatic cell damage from diabetes mellitus type 1, graft versus host disease, inflammatory bowel disease, inflamation associated with pulmonary disease, other autoimmune disease or inflammatory disease, an antiproliferative agent such as for acute or chronic mylegenous leukemia or in the prevention of premature labor secondary to intrauterine infections.

4.7.16 LEUKEMIAS

Leukemias and related disorders may be treated or prevented by administration of a therapeutic that promotes or inhibits function of the polynucleotides and/or polypeptides of the invention. Such leukemias and related disorders include but are not limited to acute leukemia, acute lymphocytic leukemia, acute myelocytic leukemia, myeloblastic, promyelocytic, myelomonocytic, monocytic, erythroleukemia, chronic leukemia, chronic myelocytic (granulocytic) leukemia and chronic lymphocytic leukemia (for a review of such disorders, see Fishman et al., 1985, Medicine, 2d Ed., J.B. Lippincott Co., Philadelphia).

4.7.17 NERVOUS SYSTEM DISORDERS

Nervous system disorders, involving cell types which can be tested for efficacy of intervention with compounds that modulate the activity of the polynucleotides and/or polypeptides of the invention, and which can be treated upon thus observing an indication of therapeutic utility, include but are not limited to nervous system injuries, and diseases

or disorders which result in either a disconnection of axons, a diminution or degeneration of neurons, or demyelination. Nervous system lesions which may be treated in a patient (including human and non-human mammalian patients) according to the invention include but are not limited to the following lesions of either the central (including spinal cord, brain) or peripheral nervous systems:

(i) traumatic lesions, including lesions caused by physical injury or associated with surgery, for example, lesions which sever a portion of the nervous system, or compression injuries;

(ii) ischemic lesions, in which a lack of oxygen in a portion of the nervous system results in neuronal injury or death, including cerebral infarction or ischemia, or spinal cord infarction or ischemia;

(iii) infectious lesions, in which a portion of the nervous system is destroyed or injured as a result of infection, for example, by an abscess or associated with infection by human immunodeficiency virus, herpes zoster, or herpes simplex virus or with Lyme disease, tuberculosis, syphilis;

(iv) degenerative lesions, in which a portion of the nervous system is destroyed or injured as a result of a degenerative process including but not limited to degeneration associated with Parkinson's disease, Alzheimer's disease, Huntington's chorea, or amyotrophic lateral sclerosis;

(v) lesions associated with nutritional diseases or disorders, in which a portion of the nervous system is destroyed or injured by a nutritional disorder or disorder of metabolism including but not limited to, vitamin B12 deficiency, folic acid deficiency, Wernicke disease, tobacco-alcohol amblyopia, Marchiafava-Bignami disease (primary degeneration of the corpus callosum), and alcoholic cerebellar degeneration;

(vi) neurological lesions associated with systemic diseases including but not limited to diabetes (diabetic neuropathy, Bell's palsy), systemic lupus erythematosus, carcinoma, or sarcoidosis;

(vii) lesions caused by toxic substances including alcohol, lead, or particular neurotoxins; and

(viii) demyelinated lesions in which a portion of the nervous system is destroyed or injured by a demyelinating disease including but not limited to multiple

sclerosis, human immunodeficiency virus-associated myelopathy, transverse myelopathy or various etiologies, progressive multifocal leukoencephalopathy, and central pontine myelinolysis.

Therapeutics which are useful according to the invention for treatment of a nervous system disorder may be selected by testing for biological activity in promoting the survival or differentiation of neurons. For example, and not by way of limitation, therapeutics which elicit any of the following effects may be useful according to the invention:

(i) increased survival time of neurons in culture;

(ii) increased sprouting of neurons in culture or in vivo;

(iii) increased production of a neuron-associated molecule in culture or *in vivo*, *e.g.*, choline acetyltransferase or acetylcholinesterase with respect to motor neurons; or

(iv) decreased symptoms of neuron dysfunction in vivo.

Such effects may be measured by any method known in the art. In preferred, non-limiting embodiments, increased survival of neurons may be measured by the method set forth in Arakawa et al. (1990, J. Neurosci. 10:3507-3515); increased sprouting of neurons may be detected by methods set forth in Pestronk et al. (1980, Exp. Neurol. 70:65-82) or Brown et al. (1981, Ann. Rev. Neurosci. 4:17-42); increased production of neuron-associated molecules may be measured by bioassay, enzymatic assay, antibody binding, Northern blot assay, *etc.*, depending on the molecule to be measured; and motor neuron dysfunction may be measured by assessing the physical manifestation of motor neuron disorder, *e.g.*, weakness, motor neuron conduction velocity, or functional disability.

In specific embodiments, motor neuron disorders that may be treated according to the invention include but are not limited to disorders such as infarction, infection, exposure to toxin, trauma, surgical damage, degenerative disease or malignancy that may affect motor neurons as well as other components of the nervous system, as well as disorders that selectively affect neurons such as amyotrophic lateral sclerosis, and including but not limited to progressive spinal muscular atrophy, progressive bulbar palsy, primary lateral sclerosis, infantile and juvenile muscular atrophy, progressive

bulbar paralysis of childhood (Fazio-Londe syndrome), poliomyelitis and the post polio syndrome, and Hereditary Motorsensory Neuropathy (Charcot-Marie-Tooth Disease).

4.7.18 OTHER ACTIVITIES

A polypeptide of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, co-factors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

4.7.19 IDENTIFICATION OF POLYMORPHISMS

The demonstration of polymorphisms makes possible the identification of such polymorphisms in human subjects and the pharmacogenetic use of this information for diagnosis and treatment. Such polymorphisms may be associated with, e.g., differential predisposition or susceptibility to various disease states (such as disorders involving

inflammation or immune response) or a differential response to drug administration, and this genetic information can be used to tailor preventive or therapeutic treatment appropriately. For example, the existence of a polymorphism associated with a predisposition to inflammation or autoimmune disease makes possible the diagnosis of this condition in humans by identifying the presence of the polymorphism.

Polymorphisms can be identified in a variety of ways known in the art which all generally involve obtaining a sample from a patient, analyzing DNA from the sample, optionally involving isolation or amplification of the DNA, and identifying the presence of the polymorphism in the DNA. For example, PCR may be used to amplify an appropriate fragment of genomic DNA which may then be sequenced. Alternatively, the DNA may be subjected to allele-specific oligonucleotide hybridization (in which appropriate oligonucleotides are hybridized to the DNA under conditions permitting detection of a single base mismatch) or to a single nucleotide extension assay (in which an oligonucleotide that hybridizes immediately adjacent to the position of the polymorphism is extended with one or more labeled nucleotides). In addition, traditional restriction fragment length polymorphism analysis (using restriction enzymes that provide differential digestion of the genomic DNA depending on the presence or absence of the polymorphism) may be performed. Arrays with nucleotide sequences of the present invention can be used to detect polymorphisms. The array can comprise modified nucleotide sequences of the present invention in order to detect the nucleotide sequences of the present invention. In the alternative, any one of the nucleotide sequences of the present invention can be placed on the array to detect changes from those sequences.

Alternatively a polymorphism resulting in a change in the amino acid sequence could also be detected by detecting a corresponding change in amino acid sequence of the protein, e.g., by an antibody specific to the variant sequence.

4.7.20 ARTHRITIS AND INFLAMMATION

The immunosuppressive effects of the compositions of the invention against rheumatoid arthritis is determined in an experimental animal model system. The experimental model system is adjuvant induced arthritis in rats, and the protocol is described by J. Holoshitz, et at., 1983, Science, 219:56, or by B. Waksman et al., 1963,

Int. Arch. Allergy Appl. Immunol., 23:129. Induction of the disease can be caused by a single injection, generally intradermally, of a suspension of killed Mycobacterium tuberculosis in complete Freund's adjuvant (CFA). The route of injection can vary, but rats may be injected at the base of the tail with an adjuvant mixture. The polypeptide is administered in phosphate buffered solution (PBS) at a dose of about 1-5 mg/kg. The control consists of administering PBS only.

The procedure for testing the effects of the test compound would consist of intradermally injecting killed Mycobacterium tuberculosis in CFA followed by immediately administering the test compound and subsequent treatment every other day until day 24. At 14, 15, 18, 20, 22, and 24 days after injection of Mycobacterium CFA, an overall arthritis score may be obtained as described by J. Holoskitz above. An analysis of the data would reveal that the test compound would have a dramatic affect on the swelling of the joints as measured by a decrease of the arthritis score.

4.8 THERAPEUTIC METHODS

The compositions (including polypeptide fragments, analogs, variants and antibodies or other binding partners or modulators including antisense polynucleotides) of the invention have numerous applications in a variety of therapeutic methods. Examples of therapeutic applications include, but are not limited to, those exemplified herein.

4.8.1 EXAMPLE

One embodiment of the invention is the administration of an effective amount of the polypeptides or other composition of the invention to individuals affected by a disease or disorder that can be modulated by regulating the peptides of the invention. While the mode of administration is not particularly important, parenteral administration is preferred. An exemplary mode of administration is to deliver an intravenous bolus. The dosage of the polypeptides or other composition of the invention will normally be determined by the prescribing physician. It is to be expected that the dosage will vary according to the age, weight, condition and response of the individual patient. Typically,

the amount of polypeptide administered per dose will be in the range of about 0.01μg/kg to 100 mg/kg of body weight, with the preferred dose being about 0.1μg/kg to 10 mg/kg of patient body weight. For parenteral administration, polypeptides of the invention will be formulated in an injectable form combined with a pharmaceutically acceptable parenteral vehicle. Such vehicles are well known in the art and examples include water, saline, Ringer's solution, dextrose solution, and solutions consisting of small amounts of the human serum albumin. The vehicle may contain minor amounts of additives that maintain the isotonicity and stability of the polypeptide or other active ingredient. The preparation of such solutions is within the skill of the art.

4.9 PHARMACEUTICAL FORMULATIONS AND ROUTES OF ADMINISTRATION

A protein or other composition of the present invention (from whatever source derived, including without limitation from recombinant and non-recombinant sources and including antibodies and other binding partners of the polypeptides of the invention) may be administered to a patient in need, by itself, or in pharmaceutical compositions where it is mixed with suitable carriers or excipient(s) at doses to treat or ameliorate a variety of disorders. Such a composition may optionally contain (in addition to protein or other active ingredient and a carrier) diluents, fillers, salts, buffers, stabilizers, solubilizers, and other materials well known in the art. The term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the effectiveness of the biological activity of the active ingredient(s). The characteristics of the carrier will depend on the route of administration. The pharmaceutical composition of the invention may also contain cytokines, lymphokines, or other hematopoietic factors such as M-CSF, GM-CSF, TNF, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IFN, TNF0, TNF1, TNF2, G-CSF, Meg-CSF, thrombopoietin, stem cell factor, and erythropoietin. In further compositions, proteins of the invention may be combined with other agents beneficial to the treatment of the disease or disorder in question. These agents include various growth factors such as epidermal growth factor (EGF),

platelet-derived growth factor (PDGF), transforming growth factors (TGF-α and TGF-β), insulin-like growth factor (IGF), as well as cytokines described herein.

The pharmaceutical composition may further contain other agents which either enhance the activity of the protein or other active ingredient or complement its activity or use in treatment. Such additional factors and/or agents may be included in the pharmaceutical composition to produce a synergistic effect with protein or other active ingredient of the invention, or to minimize side effects. Conversely, protein or other active ingredient of the present invention may be included in formulations of the particular clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti- inflammatory agent to minimize side effects of the clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent (such as IL-1Ra, IL-1 Hy1, IL-1 Hy2, anti-TNF, corticosteroids, immunosuppressive agents). A protein of the present invention may be active in multimers (e.g., heterodimers or homodimers) or complexes with itself or other proteins. As a result, pharmaceutical compositions of the invention may comprise a protein of the invention in such multimeric or complexed form.

As an alternative to being included in a pharmaceutical composition of the invention including a first protein, a second protein or a therapeutic agent may be concurrently administered with the first protein (e.g., at the same time, or at differing times provided that therapeutic concentrations of the combination of agents is achieved at the treatment site). Techniques for formulation and administration of the compounds of the instant application may be found in "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, PA, latest edition. A therapeutically effective dose further refers to that amount of the compound sufficient to result in amelioration of symptoms, *e.g.*, treatment, healing, prevention or amelioration of the relevant medical condition, or an increase in rate of treatment, healing, prevention or amelioration of such conditions. When applied to an individual active ingredient, administered alone, a therapeutically effective dose refers to that ingredient alone. When applied to a combination, a therapeutically effective dose refers to combined amounts of the active ingredients that result in the therapeutic effect, whether administered in combination, serially or simultaneously.

In practicing the method of treatment or use of the present invention, a therapeutically effective amount of protein or other active ingredient of the present invention is administered to a mammal having a condition to be treated. Protein or other active ingredient of the present invention may be administered in accordance with the method of the invention either alone or in combination with other therapies such as treatments employing cytokines, lymphokines or other hematopoietic factors. When co-administered with one or more cytokines, lymphokines or other hematopoietic factors, protein or other active ingredient of the present invention may be administered either simultaneously with the cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors, or sequentially. If administered sequentially, the attending physician will decide on the appropriate sequence of administering protein or other active ingredient of the present invention in combination with cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors.

4.9.1 ROUTES OF ADMINISTRATION

Suitable routes of administration may, for example, include oral, rectal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections. Administration of protein or other active ingredient of the present invention used in the pharmaceutical composition or to practice the method of the present invention can be carried out in a variety of conventional ways, such as oral ingestion, inhalation, topical application or cutaneous, subcutaneous, intraperitoneal, parenteral or intravenous injection. Intravenous administration to the patient is preferred.

Alternately, one may administer the compound in a local rather than systemic manner, for example, via injection of the compound directly into a arthritic joints or in fibrotic tissue, often in a depot or sustained release formulation. In order to prevent the scarring process frequently occurring as complication of glaucoma surgery, the compounds may be administered topically, for example, as eye drops. Furthermore, one may administer the drug in a targeted drug delivery system, for example, in a liposome

coated with a specific antibody, targeting, for example, arthritic or fibrotic tissue. The liposomes will be targeted to and taken up selectively by the afflicted tissue.

The polypeptides of the invention are administered by any route that delivers an effective dosage to the desired site of action. The determination of a suitable route of administration and an effective dosage for a particular indication is within the level of skill in the art. Preferably for wound treatment, one administers the therapeutic compound directly to the site. Suitable dosage ranges for the polypeptides of the invention can be extrapolated from these dosages or from similar studies in appropriate animal models. Dosages can then be adjusted as necessary by the clinician to provide maximal therapeutic benefit.

4.9.2 COMPOSITIONS/FORMULATIONS

Pharmaceutical compositions for use in accordance with the present invention thus may be formulated in a conventional manner using one or more physiologically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. These pharmaceutical compositions may be manufactured in a manner that is itself known, *e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes. Proper formulation is dependent upon the route of administration chosen. When a therapeutically effective amount of protein or other active ingredient of the present invention is administered orally, protein or other active ingredient of the present invention will be in the form of a tablet, capsule, powder, solution or elixir. When administered in tablet form, the pharmaceutical composition of the invention may additionally contain a solid carrier such as a gelatin or an adjuvant. The tablet, capsule, and powder contain from about 5 to 95% protein or other active ingredient of the present invention, and preferably from about 25 to 90% protein or other active ingredient of the present invention. When administered in liquid form, a liquid carrier such as water, petroleum, oils of animal or plant origin such as peanut oil, mineral oil, soybean oil, or sesame oil, or synthetic oils may be added. The liquid form of the pharmaceutical composition may further contain physiological saline solution, dextrose or other saccharide solution, or glycols such as ethylene glycol,

propylene glycol or polyethylene glycol. When administered in liquid form, the pharmaceutical composition contains from about 0.5 to 90% by weight of protein or other active ingredient of the present invention, and preferably from about 1 to 50% protein or other active ingredient of the present invention.

When a therapeutically effective amount of protein or other active ingredient of the present invention is administered by intravenous, cutaneous or subcutaneous injection, protein or other active ingredient of the present invention will be in the form of a pyrogen-free, parenterally acceptable aqueous solution. The preparation of such parenterally acceptable protein or other active ingredient solutions, having due regard to pH, isotonicity, stability, and the like, is within the skill in the art. A preferred pharmaceutical composition for intravenous, cutaneous, or subcutaneous injection should contain, in addition to protein or other active ingredient of the present invention, an isotonic vehicle such as Sodium Chloride Injection, Ringer's Injection, Dextrose Injection, Dextrose and Sodium Chloride Injection, Lactated Ringer's Injection, or other vehicle as known in the art. The pharmaceutical composition of the present invention may also contain stabilizers, preservatives, buffers, antioxidants, or other additives known to those of skill in the art. For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiological saline buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

For oral administration, the compounds can be formulated readily by combining the active compounds with pharmaceutically acceptable carriers well known in the art. Such carriers enable the compounds of the invention to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated. Pharmaceutical preparations for oral use can be obtained from a solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose,

hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate. Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added. All formulations for oral administration should be in dosages suitable for such administration. For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, *e.g.*, gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch. The compounds may be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, *e.g.*, in ampules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or

aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other glycerides. In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

A pharmaceutical carrier for the hydrophobic compounds of the invention is a co-solvent system comprising benzyl alcohol, a nonpolar surfactant, a water-miscible organic polymer, and an aqueous phase. The co-solvent system may be the VPD co-solvent system. VPD is a solution of 3% w/v benzyl alcohol, 8% w/v of the nonpolar surfactant polysorbate 80, and 65% w/v polyethylene glycol 300, made up to volume in absolute ethanol. The VPD co-solvent system (VPD:5W) consists of VPD diluted 1:1 with a 5% dextrose in water solution. This co-solvent system dissolves hydrophobic compounds well, and itself produces low toxicity upon systemic administration. Naturally, the proportions of a co-solvent system may be varied considerably without

destroying its solubility and toxicity characteristics. Furthermore, the identity of the co-solvent components may be varied: for example, other low-toxicity nonpolar surfactants may be used instead of polysorbate 80; the fraction size of polyethylene glycol may be varied; other biocompatible polymers may replace polyethylene glycol, e.g. polyvinyl pyrrolidone; and other sugars or polysaccharides may substitute for dextrose. Alternatively, other delivery systems for hydrophobic pharmaceutical compounds may be employed. Liposomes and emulsions are well known examples of delivery vehicles or carriers for hydrophobic drugs. Certain organic solvents such as dimethylsulfoxide also may be employed, although usually at the cost of greater toxicity. Additionally, the compounds may be delivered using a sustained-release system, such as semipermeable matrices of solid hydrophobic polymers containing the therapeutic agent. Various types of sustained-release materials have been established and are well known by those skilled in the art. Sustained-release capsules may, depending on their chemical nature, release the compounds for a few weeks up to over 100 days. Depending on the chemical nature and the biological stability of the therapeutic reagent, additional strategies for protein or other active ingredient stabilization may be employed.

The pharmaceutical compositions also may comprise suitable solid or gel phase carriers or excipients. Examples of such carriers or excipients include but are not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols. Many of the active ingredients of the invention may be provided as salts with pharmaceutically compatible counter ions. Such pharmaceutically acceptable base addition salts are those salts which retain the biological effectiveness and properties of the free acids and which are obtained by reaction with inorganic or organic bases such as sodium hydroxide, magnesium hydroxide, ammonia, trialkylamine, dialkylamine, monoalkylamine, dibasic amino acids, sodium acetate, potassium benzoate, triethanol amine and the like.

The pharmaceutical composition of the invention may be in the form of a complex of the protein(s) or other active ingredient(s) of present invention along with protein or peptide antigens. The protein and/or peptide antigen will deliver a stimulatory signal to both B and T lymphocytes. B lymphocytes will respond to antigen through their surface immunoglobulin receptor. T lymphocytes will respond to antigen through the T

cell receptor (TCR) following presentation of the antigen by MHC proteins. MHC and structurally related proteins including those encoded by class I and class II MHC genes on host cells will serve to present the peptide antigen(s) to T lymphocytes. The antigen components could also be supplied as purified MHC-peptide complexes alone or with co-stimulatory molecules that can directly signal T cells. Alternatively antibodies able to bind surface immunoglobulin and other molecules on B cells as well as antibodies able to bind the TCR and other molecules on T cells can be combined with the pharmaceutical composition of the invention.

The pharmaceutical composition of the invention may be in the form of a liposome in which protein of the present invention is combined, in addition to other pharmaceutically acceptable carriers, with amphipathic agents such as lipids which exist in aggregated form as micelles, insoluble monolayers, liquid crystals, or lamellar layers in aqueous solution. Suitable lipids for liposomal formulation include, without limitation, monoglycerides, diglycerides, sulfatides, lysolecithins, phospholipids, saponin, bile acids, and the like. Preparation of such liposomal formulations is within the level of skill in the art, as disclosed, for example, in U.S. Patent Nos. 4,235,871; 4,501,728; 4,837,028; and 4,737,323, all of which are incorporated herein by reference.

The amount of protein or other active ingredient of the present invention in the pharmaceutical composition of the present invention will depend upon the nature and severity of the condition being treated, and on the nature of prior treatments which the patient has undergone. Ultimately, the attending physician will decide the amount of protein or other active ingredient of the present invention with which to treat each individual patient. Initially, the attending physician will administer low doses of protein or other active ingredient of the present invention and observe the patient's response. Larger doses of protein or other active ingredient of the present invention may be administered until the optimal therapeutic effect is obtained for the patient, and at that point the dosage is not increased further. It is contemplated that the various pharmaceutical compositions used to practice the method of the present invention should contain about 0.01 μg to about 100 mg (preferably about 0.1 μg to about 10 mg, more preferably about 0.1 μg to about 1 mg) of protein or other active ingredient of the present invention per kg body weight. For compositions of the present invention which are

useful for bone, cartilage, tendon or ligament regeneration, the therapeutic method includes administering the composition topically, systematically, or locally as an implant or device. When administered, the therapeutic composition for use in this invention is, of course, in a pyrogen-free, physiologically acceptable form. Further, the composition may desirably be encapsulated or injected in a viscous form for delivery to the site of bone, cartilage or tissue damage. Topical administration may be suitable for wound healing and tissue repair. Therapeutically useful agents other than a protein or other active ingredient of the invention which may also optionally be included in the composition as described above, may alternatively or additionally, be administered simultaneously or sequentially with the composition in the methods of the invention. Preferably for bone and/or cartilage formation, the composition would include a matrix capable of delivering the protein-containing or other active ingredient-containing composition to the site of bone and/or cartilage damage, providing a structure for the developing bone and cartilage and optimally capable of being resorbed into the body. Such matrices may be formed of materials presently in use for other implanted medical applications.

The choice of matrix material is based on biocompatibility, biodegradability, mechanical properties, cosmetic appearance and interface properties. The particular application of the compositions will define the appropriate formulation. Potential matrices for the compositions may be biodegradable and chemically defined calcium sulfate, tricalcium phosphate, hydroxyapatite, polylactic acid, polyglycolic acid and polyanhydrides. Other potential materials are biodegradable and biologically well-defined, such as bone or dermal collagen. Further matrices are comprised of pure proteins or extracellular matrix components. Other potential matrices are nonbiodegradable and chemically defined, such as sintered hydroxyapatite, bioglass, aluminates, or other ceramics. Matrices may be comprised of combinations of any of the above mentioned types of material, such as polylactic acid and hydroxyapatite or collagen and tricalcium phosphate. The bioceramics may be altered in composition, such as in calcium-aluminate-phosphate and processing to alter pore size, particle size, particle shape, and biodegradability. Presently preferred is a 50:50 (mole weight) copolymer of lactic acid and glycolic acid in the form of porous particles having diameters ranging from 150 to 800 microns. In some applications, it will be useful to utilize a sequestering

agent, such as carboxymethyl cellulose or autologous blood clot, to prevent the protein compositions from disassociating from the matrix.

A preferred family of sequestering agents is cellulosic materials such as alkylcelluloses (including hydroxyalkylcelluloses), including methylcellulose, ethylcellulose, hydroxyethylcellulose, hydroxypropylcellulose, hydroxypropyl-methylcellulose, and carboxymethylcellulose, the most preferred being cationic salts of carboxymethylcellulose (CMC). Other preferred sequestering agents include hyaluronic acid, sodium alginate, poly(ethylene glycol), polyoxyethylene oxide, carboxyvinyl polymer and poly(vinyl alcohol). The amount of sequestering agent useful herein is 0.5-20 wt %, preferably 1-10 wt % based on total formulation weight, which represents the amount necessary to prevent desorption of the protein from the polymer matrix and to provide appropriate handling of the composition, yet not so much that the progenitor cells are prevented from infiltrating the matrix, thereby providing the protein the opportunity to assist the osteogenic activity of the progenitor cells. In further compositions, proteins or other active ingredients of the invention may be combined with other agents beneficial to the treatment of the bone and/or cartilage defect, wound, or tissue in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet derived growth factor (PDGF), transforming growth factors (TGF-α and TGF-β), and insulin-like growth factor (IGF).

The therapeutic compositions are also presently valuable for veterinary applications. Particularly domestic animals and thoroughbred horses, in addition to humans, are desired patients for such treatment with proteins or other active ingredients of the present invention. The dosage regimen of a protein-containing pharmaceutical composition to be used in tissue regeneration will be determined by the attending physician considering various factors which modify the action of the proteins, *e.g.*, amount of tissue weight desired to be formed, the site of damage, the condition of the damaged tissue, the size of a wound, type of damaged tissue (*e.g.*, bone), the patient's age, sex, and diet, the severity of any infection, time of administration and other clinical factors. The dosage may vary with the type of matrix used in the reconstitution and with inclusion of other proteins in the pharmaceutical composition. For example, the addition of other known growth factors, such as IGF I (insulin like growth factor I), to the final

composition, may also effect the dosage. Progress can be monitored by periodic assessment of tissue/bone growth and/or repair, for example, X-rays, histomorphometric determinations and tetracycline labeling.

Polynucleotides of the present invention can also be used for gene therapy. Such polynucleotides can be introduced either in vivo or ex vivo into cells for expression in a mammalian subject. Polynucleotides of the invention may also be administered by other known methods for introduction of nucleic acid into a cell or organism (including, without limitation, in the form of viral vectors or naked DNA). Cells may also be cultured ex vivo in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced in vivo for therapeutic purposes.

4.9.3 EFFECTIVE DOSAGE

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve its intended purpose. More specifically, a therapeutically effective amount means an amount effective to prevent development of or to alleviate the existing symptoms of the subject being treated. Determination of the effective amount is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from appropriate in vitro assays. For example, a dose can be formulated in animal models to achieve a circulating concentration range that can be used to more accurately determine useful doses in humans. For example, a dose can be formulated in animal models to achieve a circulating concentration range that includes the IC_{50} as determined in cell culture (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of the protein's biological activity). Such information can be used to more accurately determine useful doses in humans.

A therapeutically effective dose refers to that amount of the compound that results in amelioration of symptoms or a prolongation of survival in a patient. Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical

procedures in cell cultures or experimental animals, *e.g.*, for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio between LD_{50} and ED_{50}. Compounds which exhibit high therapeutic indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. See, *e.g.*, Fingl et al., 1975, in "The Pharmacological Basis of Therapeutics", Ch. 1 p.1. Dosage amount and interval may be adjusted individually to provide plasma levels of the active moiety which are sufficient to maintain the desired effects, or minimal effective concentration (MEC). The MEC will vary for each compound but can be estimated from *in vitro* data. Dosages necessary to achieve the MEC will depend on individual characteristics and route of administration. However, HPLC assays or bioassays can be used to determine plasma concentrations.

Dosage intervals can also be determined using MEC value. Compounds should be administered using a regimen which maintains plasma levels above the MEC for 10-90% of the time, preferably between 30-90% and most preferably between 50-90%. In cases of local administration or selective uptake, the effective local concentration of the drug may not be related to plasma concentration.

An exemplary dosage regimen for polypeptides or other compositions of the invention will be in the range of about 0.01 μg/kg to 100 mg/kg of body weight daily, with the preferred dose being about 0.1 μg/kg to 25 mg/kg of patient body weight daily, varying in adults and children. Dosing may be once daily, or equivalent doses may be delivered at longer or shorter intervals.

The amount of composition administered will, of course, be dependent on the subject being treated, on the subject's age and weight, the severity of the affliction, the manner of administration and the judgment of the prescribing physician.

4.9.4 PACKAGING

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may, for example, comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration. Compositions comprising a compound of the invention formulated in a compatible pharmaceutical carrier may also be prepared, placed in an appropriate container, and labeled for treatment of an indicated condition.

4.10 ANTIBODIES

Another aspect of the invention is an antibody that specifically binds the polypeptide of the invention. Such antibodies include monoclonal and polyclonal antibodies, single chain antibodies, chimeric antibodies, bifunctional/bispecific antibodies, humanized antibodies, human antibodies, and complementary determining region (CDR)-grafted antibodies, including compounds which include CDR and/or antigen-binding sequences, which specifically recognize a polypeptide of the invention. Preferred antibodies of the invention are human antibodies which are produced and identified according to methods described in WO93/11236, published June 20, 1993, which is incorporated herein by reference in its entirety. Antibody fragments, including Fab, Fab′, $F(ab')_2$, and F_v, are also provided by the invention. The term "specific for" indicates that the variable regions of the antibodies of the invention recognize and bind polypeptides of the invention exclusively (*i.e.*, able to distinguish the polypeptide of the invention from other similar polypeptides despite sequence identity, homology, or similarity found in the family of polypeptides), but may also interact with other proteins (for example, *S. aureus* protein A or other antibodies in ELISA techniques) through interactions with sequences outside the variable region of the antibodies, and in particular, in the constant region of the molecule. Screening assays to determine binding specificity of an antibody of the invention are well known and routinely practiced in the art. For a comprehensive discussion of such assays, see Harlow et al. (Eds), Antibodies A Laboratory Manual; Cold Spring Harbor Laboratory; Cold Spring Harbor , NY (1988),

Chapter 6. Antibodies that recognize and bind fragments of the polypeptides of the invention are also contemplated, provided that the antibodies are first and foremost specific for, as defined above, full length polypeptides of the invention. As with antibodies that are specific for full length polypeptides of the invention, antibodies of the invention that recognize fragments are those which can distinguish polypeptides from the same family of polypeptides despite inherent sequence identity, homology, or similarity found in the family of proteins. Antibodies of the invention can be produced using any method well known and routinely practiced in the art.

Non-human antibodies may be humanized by any methods known in the art. In one method, the non-human CDRs are inserted into a human antibody or consensus antibody framework sequence. Further changes can then be introduced into the antibody framework to modulate affinity or immunogenicity.

Antibodies of the invention are useful for, for example, therapeutic purposes (by modulating activity of a polypeptide of the invention), diagnostic purposes to detect or quantitate a polypeptide of the invention, as well as purification of a polypeptide of the invention. Kits comprising an antibody of the invention for any of the purposes described herein are also comprehended. In general, a kit of the invention also includes a control antigen for which the antibody is immunospecific. The invention further provides a hybridoma that produces an antibody according to the invention. Antibodies of the invention are useful for detection and/or purification of the polypeptides of the invention.

Polypeptides of the invention may also be used to immunize animals to obtain polyclonal and monoclonal antibodies which specifically react with the protein. Such antibodies may be obtained using either the entire protein or fragments thereof as an immunogen. The peptide immunogens additionally may contain a cysteine residue at the carboxyl terminus, and are conjugated to a hapten such as keyhole limpet hemocyanin (KLH). Methods for synthesizing such peptides are known in the art, for example, as in R. P. Merrifield, J. Amer. Chem. Soc. 85, 2149-2154 (1963); J. L. Krstenansky, et al., FEBS Lett. 211, 10 (1987).

Monoclonal antibodies binding to the protein of the invention may be useful diagnostic agents for the immunodetection of the protein. Neutralizing monoclonal antibodies binding to the protein may also be useful therapeutics for both conditions

associated with the protein and also in the treatment of some forms of cancer where abnormal expression of the protein is involved. In the case of cancerous cells or leukemic cells, neutralizing monoclonal antibodies against the protein may be useful in detecting and preventing the metastatic spread of the cancerous cells, which may be mediated by the protein. In general, techniques for preparing polyclonal and monoclonal antibodies as well as hybridomas capable of producing the desired antibody are well known in the art (Campbell, A.M., Monoclonal Antibodies Technology: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1984); St. Groth et al., J. Immunol. 35:1-21 (1990); Kohler and Milstein, Nature 256:495-497 (1975)), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., Immunology Today 4:72 (1983); Cole et al., in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc. (1985), pp. 77-96).

Any animal (mouse, rabbit, *etc.*) which is known to produce antibodies can be immunized with a peptide or polypeptide of the invention. Methods for immunization are well known in the art. Such methods include subcutaneous or intraperitoneal injection of the polypeptide. One skilled in the art will recognize that the amount of the protein encoded by the ORF of the present invention used for immunization will vary based on the animal which is immunized, the antigenicity of the peptide and the site of injection. The protein that is used as an immunogen may be modified or administered in an adjuvant in order to increase the protein's antigenicity. Methods of increasing the antigenicity of a protein are well known in the art and include, but are not limited to, coupling the antigen with a heterologous protein (such as globulin or β-galactosidase) or through the inclusion of an adjuvant during immunization.

For monoclonal antibodies, spleen cells from the immunized animals are removed, fused with myeloma cells, such as SP2/0-Ag14 myeloma cells, and allowed to become monoclonal antibody producing hybridoma cells. Any one of a number of methods well known in the art can be used to identify the hybridoma cell which produces an antibody with the desired characteristics. These include screening the hybridomas with an ELISA assay, Western blot analysis, or radioimmunoassay (Lutz et al., Exp. Cell Research. 175:109-124 (1988)). Hybridomas secreting the desired antibodies are cloned and the class and subclass is determined using procedures known in the art (Campbell,

A.M., Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1984)). Techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce single chain antibodies to proteins of the present invention.

For polyclonal antibodies, antibody-containing antiserum is isolated from the immunized animal and is screened for the presence of antibodies with the desired specificity using one of the above-described procedures. The present invention further provides the above- described antibodies in delectably labeled form. Antibodies can be delectably labeled through the use of radioisotopes, affinity labels (such as biotin, avidin, etc.), enzymatic labels (such as horseradish peroxidase, alkaline phosphatase, etc.) fluorescent labels (such as FITC or rhodamine, etc.), paramagnetic atoms, etc. Procedures for accomplishing such labeling are well-known in the art, for example, see (Sternberger, L.A. et al., J. Histochem. Cytochem. 18:315 (1970); Bayer, E.A. et al., Meth. Enzym. 62:308 (1979); Engval, E. et al., Immunol. 109:129 (1972); Goding, J.W. J. Immunol. Meth. 13:215 (1976)).

The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays to identify cells or tissues in which a fragment of the polypeptide of interest is expressed. The antibodies may also be used directly in therapies or other diagnostics. The present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and Sepharose®, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir, D.M. et al., "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W.D. et al., Meth. Enzym. 34 Academic Press, N.Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays as well as for immuno-affinity purification of the proteins of the present invention.

4.11 COMPUTER READABLE SEQUENCES

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (*e.g.* text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

By providing any of the nucleotide sequences SEQ ID NOs: 1 - 35 or a representative fragment thereof; or a nucleotide sequence at least 95% identical to any of the nucleotide sequences of the SEQ ID NOs: 1 - 35 in computer readable form, a skilled artisan can routinely access the sequence information for a variety of purposes. Computer software is publicly available which allows a skilled artisan to access sequence

information provided in a computer readable medium. The examples which follow demonstrate how software which implements the BLAST (Altschul et al., J. Mol. Biol. 215:403-410 (1990)) and BLAZE (Brutlag et al., Comp. Chem. 17:203-207 (1993)) search algorithms on a Sybase system is used to identify open reading frames (ORFs) within a nucleic acid sequence. Such ORFs may be protein encoding fragments and may be useful in producing commercially important proteins such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems are suitable for use in the present invention. As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means. As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of a known sequence which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, Smith-Waterman, MacPattern (EMBL), BLASTN and BLASTA (NPOLYPEPTIDEIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based

systems. As used herein, a "target sequence" can be any nucleic acid or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 300 amino acids, more preferably from about 30 to 100 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

4.12 TRIPLE HELIX FORMATION

In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Polynucleotides suitable for use in these methods are preferably 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 15241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Olmno, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide.

4.13 DIAGNOSTIC ASSAYS AND KITS

The present invention further provides methods to identify the presence or expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using a nucleic acid probe or antibodies of the present invention, optionally conjugated or otherwise associated with a suitable label.

In general, methods for detecting a polynucleotide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polynucleotide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polynucleotide of the invention is detected in the sample. Such methods can also comprise contacting a sample under stringent hybridization conditions with nucleic acid primers that anneal to a polynucleotide of the invention under such conditions, and amplifying annealed polynucleotides, so that if a polynucleotide is amplified, a polynucleotide of the invention is detected in the sample.

In general, methods for detecting a polypeptide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polypeptide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polypeptide of the invention is detected in the sample.

In detail, such methods comprise incubating a test sample with one or more of the antibodies or one or more of the nucleic acid probes of the present invention and assaying for binding of the nucleic acid probes or antibodies to components within the test sample.

Conditions for incubating a nucleic acid probe or antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid probe or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the nucleic acid probes or antibodies of the present invention. Examples of such assays can be found in Chard, T., An Introduction to Radioimmunoassay and Related Techniques, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G.R. et al., Techniques in Immunocytochemistry, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., Practice and Theory of immunoassays:

Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1985). The test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention. Specifically, the invention provides a compartment kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the probes or antibodies of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound probe or antibody.

In detail, a compartment kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound antibody or probe. Types of detection reagents include labeled nucleic acid probes, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the enzymatic, or antibody binding reagents which are capable of reacting with the labeled antibody. One skilled in the art will readily recognize that the disclosed probes and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

4.14 MEDICAL IMAGING

The novel polypeptides and binding partners of the invention are useful in medical imaging of sites expressing the molecules of the invention (e.g., where the polypeptide of the invention is involved in the immune response, for imaging sites of inflammation or infection). See, e.g., Kunkel et al., U.S. Pat. NO. 5,413,778. Such methods involve chemical attachment of a labeling or imaging agent, administration of the labeled polypeptide to a subject in a pharmaceutically acceptable carrier, and imaging the labeled polypeptide *in vivo* at the target site.

4.15 SCREENING ASSAYS

Using the isolated proteins and polynucleotides of the invention, the present invention further provides methods of obtaining and identifying agents which bind to a polypeptide encoded by an ORF corresponding to any of the nucleotide sequences set forth in the SEQ ID NOs: 1 - 35, or bind to a specific domain of the polypeptide encoded by the nucleic acid. In detail, said method comprises the steps of:

(a) contacting an agent with an isolated protein encoded by an ORF of the present invention, or nucleic acid of the invention; and

(b) determining whether the agent binds to said protein or said nucleic acid.

In general, therefore, such methods for identifying compounds that bind to a polynucleotide of the invention can comprise contacting a compound with a polynucleotide of the invention for a time sufficient to form a polynucleotide/compound complex, and detecting the complex, so that if a polynucleotide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Likewise, in general, therefore, such methods for identifying compounds that bind to a polypeptide of the invention can comprise contacting a compound with a polypeptide of the invention for a time sufficient to form a polypeptide/compound complex, and detecting the complex, so that if a polypeptide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Methods for identifying compounds that bind to a polypeptide of the invention can also comprise contacting a compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the complex drives

expression of a receptor gene sequence in the cell, and detecting the complex by detecting reporter gene sequence expression, so that if a polypeptide/compound complex is detected, a compound that binds a polypeptide of the invention is identified.

Compounds identified via such methods can include compounds which modulate the activity of a polypeptide of the invention (that is, increase or decrease its activity, relative to activity observed in the absence of the compound). Alternatively, compounds identified via such methods can include compounds which modulate the expression of a polynucleotide of the invention (that is, increase or decrease expression relative to expression levels observed in the absence of the compound). Compounds, such as compounds identified via the methods of the invention, can be tested using standard assays well known to those of skill in the art for their ability to modulate activity/expression.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention. Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed" when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like, capable of binding to a specific peptide sequence, in order to generate rationally designed antipeptide peptides, for example see Hurby et al., Application of Synthetic Peptides: Antisense Peptides," In Synthetic Peptides, A User's Guide, W.H. Freeman, NY (1992), pp. 289-307, and Kaspczak et al., Biochemistry 28:9230-8 (1989), or pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a

skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control. One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix formation by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

Agents suitable for use in these methods preferably contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide and other DNA binding agents.

Agents which bind to a protein encoded by one of the ORFs of the present invention can be used as a diagnostic agent. Agents which bind to a protein encoded by one of the ORFs of the present invention can be formulated using known techniques to generate a pharmaceutical composition.

4.16 USE OF NUCLEIC ACIDS AS PROBES

Another aspect of the subject invention is to provide for polypeptide-specific nucleic acid hybridization probes capable of hybridizing with naturally occurring nucleotide sequences. The hybridization probes of the subject invention may be derived from any of the nucleotide sequences SEQ ID NOs: 1 - 35. Because the corresponding gene is only expressed in a limited number of tissues, a hybridization probe derived from of any of the nucleotide sequences SEQ ID NOs: 1 - 35 can be used as an indicator of the presence of RNA of cell type of such a tissue in a sample.

Any suitable hybridization technique can be employed, such as, for example, in situ hybridization. PCR as described in US Patents Nos. 4,683,195 and 4,965,188 provides additional uses for oligonucleotides based upon the nucleotide sequences. Such probes used in PCR may be of recombinant origin, may be chemically synthesized, or a mixture of both. The probe will comprise a discrete nucleotide sequence for the detection of identical sequences or a degenerate pool of possible sequences for identification of closely related genomic sequences.

Other means for producing specific hybridization probes for nucleic acids include the cloning of nucleic acid sequences into vectors for the production of mRNA probes. Such vectors are known in the art and are commercially available and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerase as T7 or SP6 RNA polymerase and the appropriate radioactively labeled nucleotides. The nucleotide sequences may be used to construct hybridization probes for mapping their respective genomic sequences. The nucleotide sequence provided herein may be mapped to a chromosome or specific regions of a chromosome using well known genetic and/or chromosomal mapping techniques. These techniques include in situ hybridization, linkage analysis against known chromosomal markers, hybridization screening with libraries or flow-sorted chromosomal preparations specific to known chromosomes, and the like. The technique of fluorescent in situ hybridization of chromosome spreads has been described, among other places, in Verma et al (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York NY.

Fluorescent *in situ* hybridization of chromosomal preparations and other physical chromosome mapping techniques may be correlated with additional genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of a nucleic acid on a physical chromosomal map and a specific disease (or predisposition to a specific disease) may help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier or affected individuals.

4.17 PREPARATION OF SUPPORT BOUND OLIGONUCLEOTIDES

Oligonucleotides, i.e., small nucleic acid segments, may be readily prepared by, for example, directly synthesizing the oligonucleotide by chemical means, as is commonly practiced using an automated oligonucleotide synthesizer.

Support bound oligonucleotides may be prepared by any of the methods known to those of skill in the art using any suitable support such as glass, polystyrene or Teflon. One strategy is to precisely spot oligonucleotides synthesized by standard synthesizers. Immobilization can be achieved using passive adsorption (Inouye & Hondo, (1990) J. Clin. Microbiol. 28(6) 1469-72); using UV light (Nagata *et al.,* 1985; Dahlen *et al.,* 1987; Morrissey & Collins, (1989) Mol. Cell Probes 3(2) 189-207) or by covalent binding of base modified DNA (Keller *et al.,* 1988; 1989); all references being specifically incorporated herein.

Another strategy that may be employed is the use of the strong biotin-streptavidin interaction as a linker. For example, Broude *et al.* (1994) Proc. Natl. Acad. Sci. USA 91(8) 3072-6, describe the use of biotinylated probes, although these are duplex probes, that are immobilized on streptavidin-coated magnetic beads. Streptavidin-coated beads may be purchased from Dynal, Oslo. Of course, this same linking chemistry is applicable to coating any surface with streptavidin. Biotinylated probes may be purchased from various sources, such as, e.g., Operon Technologies (Alameda, CA).

Nunc Laboratories (Naperville, IL) is also selling suitable material that could be used. Nunc Laboratories have developed a method by which DNA can be covalently bound to the microwell surface termed Covalink NH. CovaLink NH is a polystyrene surface grafted with secondary amino groups (>NH) that serve as bridge-heads for further covalent coupling. CovaLink Modules may be purchased from Nunc Laboratories. DNA molecules may be bound to CovaLink exclusively at the 5'-end by a phosphoramidate bond, allowing immobilization of more than 1 pmol of DNA (Rasmussen *et al.,* (1991) Anal. Biochem. 198(1) 138-42).

The use of CovaLink NH strips for covalent binding of DNA molecules at the 5'-end has been described (Rasmussen et al., (1991). In this technology, a phosphoramidate bond is employed (Chu et al., (1983) Nucleic Acids Res. 11(8) 6513-29). This is beneficial as immobilization using only a single covalent bond is preferred. The phosphoramidate bond

joins the DNA to the CovaLink NH secondary amino groups that are positioned at the end of spacer arms covalently grafted onto the polystyrene surface through a 2 nm long spacer arm. To link an oligonucleotide to CovaLink NH via an phosphoramidate bond, the oligonucleotide terminus must have a 5'-end phosphate group. It is, perhaps, even possible for biotin to be covalently bound to CovaLink and then streptavidin used to bind the probes.

More specifically, the linkage method includes dissolving DNA in water (7.5 ng/ul) and denaturing for 10 min. at 95°C and cooling on ice for 10 min. Ice-cold 0.1 M 1-methylimidazole, pH 7.0 (1-$MeIm_7$), is then added to a final concentration of 10 mM 1-$MeIm_7$. A ss DNA solution is then dispensed into CovaLink NH strips (75 ul/well) standing on ice.

Carbodiimide 0.2 M 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC), dissolved in 10 mM 1-$MeIm_7$, is made fresh and 25 ul added per well. The strips are incubated for 5 hours at 50°C. After incubation the strips are washed using, e.g., Nunc-Immuno Wash; first the wells are washed 3 times, then they are soaked with washing solution for 5 min., and finally they are washed 3 times (where in the washing solution is 0.4 N NaOH, 0.25% SDS heated to 50°C).

It is contemplated that a further suitable method for use with the present invention is that described in PCT Patent Application WO 90/03382 (Southern & Maskos), incorporated herein by reference. This method of preparing an oligonucleotide bound to a support involves attaching a nucleoside 3'-reagent through the phosphate group by a covalent phosphodiester link to aliphatic hydroxyl groups carried by the support. The oligonucleotide is then synthesized on the supported nucleoside and protecting groups removed from the synthetic oligonucleotide chain under standard conditions that do not cleave the oligonucleotide from the support. Suitable reagents include nucleoside phosphoramidite and nucleoside hydrogen phosphorate.

An on-chip strategy for the preparation of DNA probe for the preparation of DNA probe arrays may be employed. For example, addressable laser-activated photodeprotection may be employed in the chemical synthesis of oligonucleotides directly on a glass surface, as described by Fodor *et al.* (1991) Science 251(4995) 767-73, incorporated herein by reference. Probes may also be immobilized on nylon supports as described by Van Ness *et al.* (1991) Nucleic Acids Res. 19(12) 3345-50; or linked to Teflon using the method of

Duncan & Cavalier (1988) Anal. Biochem. 169(1) 104-8; all references being specifically incorporated herein.

To link an oligonucleotide to a nylon support, as described by Van Ness *et al.* (1991), requires activation of the nylon surface via alkylation and selective activation of the 5'-amine of oligonucleotides with cyanuric chloride.

One particular way to prepare support bound oligonucleotides is to utilize the light-generated synthesis described by Pease *et al.,* (1994) PNAS USA 91(11) 5022-6, incorporated herein by reference). These authors used current photolithographic techniques to generate arrays of immobilized oligonucleotide probes (DNA chips). These methods, in which light is used to direct the synthesis of oligonucleotide probes in high-density, miniaturized arrays, utilize photolabile 5'-protected *N*-acyl-deoxynucleoside phosphoramidites, surface linker chemistry and versatile combinatorial synthesis strategies. A matrix of 256 spatially defined oligonucleotide probes may be generated in this manner.

4.18 PREPARATION OF NUCLEIC ACID FRAGMENTS

The nucleic acids may be obtained from any appropriate source, such as cDNAs, genomic DNA, chromosomal DNA, microdissected chromosome bands, cosmid or YAC inserts, and RNA, including mRNA without any amplification steps. For example, Sambrook *et al.* (1989) describes three protocols for the isolation of high molecular weight DNA from mammalian cells (p. 9.14-9.23).

DNA fragments may be prepared as clones in M13, plasmid or lambda vectors and/or prepared directly from genomic DNA or cDNA by PCR or other amplification methods. Samples may be prepared or dispensed in multiwell plates. About 100-1000 ng of DNA samples may be prepared in 2-500 ml of final volume.

The nucleic acids would then be fragmented by any of the methods known to those of skill in the art including, for example, using restriction enzymes as described at 9.24-9.28 of Sambrook *et al.* (1989), shearing by ultrasound and NaOH treatment.

Low pressure shearing is also appropriate, as described by Schriefer *et al.* (1990) Nucleic Acids Res. 18(24) 7455-6, incorporated herein by reference). In this method, DNA samples are passed through a small French pressure cell at a variety of low to intermediate pressures. A lever device allows controlled application of low to intermediate pressures to

the cell. The results of these studies indicate that low-pressure shearing is a useful alternative to sonic and enzymatic DNA fragmentation methods.

One particularly suitable way for fragmenting DNA is contemplated to be that using the two base recognition endonuclease, *Cvi*JI, described by Fitzgerald *et al.* (1992) Nucleic Acids Res. 20(14) 3753-62. These authors described an approach for the rapid fragmentation and fractionation of DNA into particular sizes that they contemplated to be suitable for shotgun cloning and sequencing.

The restriction endonuclease *Cvi*JI normally cleaves the recognition sequence PuGCPy between the G and C to leave blunt ends. Atypical reaction conditions, which alter the specificity of this enzyme (*Cvi*JI**), yield a quasi-random distribution of DNA fragments form the small molecule pUC19 (2688 base pairs). Fitzgerald *et al.* (1992) quantitatively evaluated the randomness of this fragmentation strategy, using a *Cvi*JI** digest of pUC19 that was size fractionated by a rapid gel filtration method and directly ligated, without end repair, to a lac Z minus M13 cloning vector. Sequence analysis of 76 clones showed that *Cvi*JI** restricts pyGCPy and PuGCPu, in addition to PuGCPy sites, and that new sequence data is accumulated at a rate consistent with random fragmentation.

As reported in the literature, advantages of this approach compared to sonication and agarose gel fractionation include: smaller amounts of DNA are required (0.2-0.5 ug instead of 2-5 ug); and fewer steps are involved (no preligation, end repair, chemical extraction, or agarose gel electrophoresis and elution are needed

Irrespective of the manner in which the nucleic acid fragments are obtained or prepared, it is important to denature the DNA to give single stranded pieces available for hybridization. This is achieved by incubating the DNA solution for 2-5 minutes at 80-90°C. The solution is then cooled quickly to 2°C to prevent renaturation of the DNA fragments before they are contacted with the chip. Phosphate groups must also be removed from genomic DNA by methods known in the art.

4.19 PREPARATION OF DNA ARRAYS

Arrays may be prepared by spotting DNA samples on a support such as a nylon membrane. Spotting may be performed by using arrays of metal pins (the positions of which correspond to an array of wells in a microtiter plate) to repeated by transfer of about 20 nl of a DNA solution to a nylon membrane. By offset printing, a density of dots higher than the

density of the wells is achieved. One to 25 dots may be accommodated in 1 mm^2, depending on the type of label used. By avoiding spotting in some preselected number of rows and columns, separate subsets (subarrays) may be formed. Samples in one subarray may be the same genomic segment of DNA (or the same gene) from different individuals, or may be different, overlapped genomic clones. Each of the subarrays may represent replica spotting of the same samples. In one example, a selected gene segment may be amplified from 64 patients. For each patient, the amplified gene segment may be in one 96-well plate (all 96 wells containing the same sample). A plate for each of the 64 patients is prepared. By using a 96-pin device, all samples may be spotted on one 8 x 12 cm membrane. Subarrays may contain 64 samples, one from each patient. Where the 96 subarrays are identical, the dot span may be 1 mm^2 and there may be a 1 mm space between subarrays.

Another approach is to use membranes or plates (available from NUNC, Naperville, Illinois) which may be partitioned by physical spacers e.g. a plastic grid molded over the membrane, the grid being similar to the sort of membrane applied to the bottom of multiwell plates, or hydrophobic strips. A fixed physical spacer is not preferred for imaging by exposure to flat phosphor-storage screens or x-ray films.

The present invention is illustrated in the following examples. Upon consideration of the present disclosure, one of skill in the art will appreciate that many other embodiments and variations may be made in the scope of the present invention. Accordingly, it is intended that the broader aspects of the present invention not be limited to the disclosure of the following examples. The present invention is not to be limited in scope by the exemplified embodiments which are intended as illustrations of single aspects of the invention, and compositions and methods which are functionally equivalent are within the scope of the invention. Indeed, numerous modifications and variations in the practice of the invention are expected to occur to those skilled in the art upon consideration of the present preferred embodiments. Consequently, the only limitations which should be placed upon the scope of the invention are those which appear in the appended claims.

All references cited within the body of the instant specification are hereby incorporated by reference in their entirety.

5.0 EXAMPLES

5.1 EXAMPLE 1

Novel Nucleic Acid Sequences Obtained From Various Libraries

A plurality of novel nucleic acids were obtained from cDNA libraries prepared from various human tissues and in some cases isolated from a genomic library derived from human chromosome using standard PCR, SBH sequence signature analysis and Sanger sequencing techniques. The inserts of the library were amplified with PCR using primers specific for the vector sequences which flank the inserts. Clones from cDNA libraries were spotted on nylon membrane filters and screened with oligonucleotide probes (e.g., 7-mers) to obtain signature sequences. The clones were clustered into groups of similar or identical sequences. Representative clones were selected for sequencing.

In some cases, the 5' sequence of the amplified inserts was then deduced using a typical Sanger sequencing protocol. PCR products were purified and subjected to fluorescent dye terminator cycle sequencing. Single pass gel sequencing was done using a 377 Applied Biosystems (ABI) sequencer to obtain the novel nucleic acid sequences. In some cases RACE (Random Amplification of cDNA Ends) was performed to further extend the sequence in the 5' direction.

5.2 EXAMPLE 2

Novel Nucleic Acids

The novel nucleic acids of the present invention of the invention were assembled from sequences that were obtained from a cDNA library by methods described in Example 1 above, and in some cases sequences obtained from one or more public databases. The nucleic acids were assembled using an EST sequence as a seed. Then a recursive algorithm was used to extend the seed EST into an extended assemblage, by pulling additional sequences from different databases (i.e., Hyseq's database containing EST sequences, dbEST version 114, gb pri 114, and UniGene version 101) that belong to this assemblage. The algorithm terminated when there was no additional sequences from the above databases that would extend the assemblage. Inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%.

Using PHRAP (Univ. of Washington) or CAP4 (Paracel), a full length gene cDNA sequence and its corresponding protein sequence were generated from the assemblage. Any frame shifts and incorrect stop codons were corrected by hand editing. During editing, the sequence was checked using FASTY and/or BLAST against Genbank (i.e., dbEST version 118, gb pri 118, UniGene version 118, Genepet release 118). Other computer programs which may have been used in the editing process were phredPhrap and Consed (University of Washington) and ed-ready, ed-ext and gc-zip-2 (Hyseq, Inc.). The full-length nucleotide and amino acid sequences, including splice variants resulting from these procedures are shown in the Sequence Listing as SEQ ID NOS: 1- 35.

Table 1 shows the various tissue sources of SEQ ID NO: 1-35.

The homology for SEQ ID NO: 1-35 were obtained by a BLASTP version 2.0al 19MP-WashU search against Genpept release 118, using BLAST algorithm. The results showed homologues for SEQ ID NO: 1-35 from Genpept. The homologues with identifiable functions for SEQ ID NO: 1-35 are shown in Table 2 below.

Using eMatrix software package (Stanford University, Stanford, CA) (Wu et al., J. Comp. Biol., Vol. 6 pp. 219-235 (1999) herein incorporated by reference), all the sequences were examined to determine whether they had identifiable signature regions. Table 3 shows the signature region found in the indicated polypeptide sequences, the description of the signature, the eMatrix p-value(s) and the position(s) of the signature within the polypeptide sequence.

Using the pFam software program (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1) pp. 320-322 (1998) herein incorporated by reference) all the polypeptide sequences were examined for domains with homology to certain peptide domains. Table 4 shows the name of the domain found, the description, the p-value and the pFam score for the identified domain within the sequence.

The nucleotide sequence within the sequences that codes for signal peptide sequences and their cleavage sites can be determine from using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark). The process for identifying prokaryotic and eukaryotic signal peptides and their cleavage sites are also disclosed by Henrik Nielson, Jacob Engelbrecht, Soren Brunak, and Gunnar von Heijne in the publication " Identification of prokaryotic and

eukaryotic signal peptides and prediction of their cleavage sites" Protein Engineering, Vol. 10, no. 1, pp. 1-6 (1997), incorporated herein by reference. A maximum S score and a mean S score, as described in the Nielson et as reference, was obtained for the polypeptide sequences. Table 5 shows the position of the signal peptide in each of the polypeptides and the maximum score and mean score associated with that signal peptide.

TABLE 1

TISSUE ORIGIN	RNA SOURCE	HYSEQ LIBRARY NAME	SEQ ID NOS:
adult brain	GIBCO	AB3001	1 5 11 19 23-24 35
adult brain	GIBCO	ABD003	3 8 11 15 18 20 23-24 26 34-35
adult brain	Clontech	ABR001	5-7 11 19
adult brain	Clontech	ABR006	6 10 12 15 18 20 27
adult brain	Clontech	ABR008	1-4 6-11 14 16 18-19 21-22 26-28 30-31 35
adult brain	BioChain	ABR012	10 35
adult brain	Invitrogen	ABR014	10
adult brain	Invitrogen	ABT004	2 6 16 21-22 24
cultured preadipocytes	Strategene	ADP001	10 24 26 31 33
adrenal gland	Clontech	ADR002	5 7 11 24 28
adult heart	GIBCO	AHR001	1 3 5-8 10-11 14-16 25 33 35
adult kidney	GIBCO	AKD001	3 5 8 11 15-16 21-22 24 28 30 33
adult kidney	Invitrogen.	AKT002	8 10 12 15 18 21-22 26 28
adult lung	GIBCO	ALG001	10 33 35
young liver	GIBCO	ALV001	2 14 17 21-22 24 26
adult liver	Invitrogen	ALV002	17-18 24 28
adult liver	Clontech	ALV003	17
adult ovary	Invitrogen	AOV001	1-2 5-6 8 10-11 14 16 18-22 24 26 28 30 33 35
adult placenta	Clontech	APL001	6
placenta	Invitrogen	APL002	10 24 35
adult spleen	GIBCO	ASP001	2 7 10-11 18 24-25 28 33
testis	GIBCO	ATS001	21-22 26 28 30
adult bladder	Invitrogen	BLD001	10
bone marrow	Clontech	BMD001	1 11 21-22 31 33 35
bone marrow	Clontech	BMD002	1 10-11 14 25-26 28
adult colon	Invitrogen	CLN001	5 11 16
Mixture of 16 tissues - mRNAs*	Various Vendors*	CTL016	1
Mixture of 16 tissues - mRNAs*	Various Vendors*	CTL021	17
Mixture of 16 tissues - mRNAs*	Various Vendors*	CTL028	10
adult cervix	BioChain	CVX001	1-2 10-11 14 18 28 33 35
endothelial cells	Strategene	EDT001	6 8 10-11 21-22 24 26 33 35
fetal brain	Clontech	FBR004	2 20 26 30
fetal brain	Clontech	FBR006	2 6 8-9 11 16 18 21-22 27 30

* The 16 tissue-mRNAs and their vendor source, are as follows: 1) Normal adult brain mRNA (Invitrogen), 2) normal adult kidney mRNA (Invitrogen), 3) normal adult liver mRNA (Invitrogen), 4) normal fetal brain mRNA (Invitrogen), 5) normal fetal kidney mRNA (Invitrogen), 6) normal fetal liver mRNA (Invitrogen), 7) normal fetal skin mRNA (Invitrogen), 8) human adrenal gland mRNA (Clontech), 9) human bone marrow mRNA (Clontech), 10) human leukemia lymphablastic mRNA (Clontech), 11) human thymus mRNA (Clontech), 12) human lymph node mRNA (Clontech), 13) human spinal cord mRNA (Clontech), 14) human thyroid mRNA (Clontech), 15) human esophagus mRNA (BioChain), 16) human conceptional umbilical cord mRNA (BioChain).

TABLE 1

TISSUE ORIGIN	RNA SOURCE	HYSEQ LIBRARY NAME	SEQ ID NOS:
			35
fetal brain	Clontech	FBRs03	21-22
fetal brain	Invitrogen	FBT002	2 8-10 18-19 24 26 30
fetal kidney	Clontech	FKD001	7 28
fetal lung	Clontech	FLG001	10
fetal lung	Invitrogen	FLG003	10 24 26
fetal liver-spleen	Columbia University	FLS001	2 6-8 10-14 17 21-22 24 26 28 31 34-35
fetal liver-spleen	Columbia University	FLS002	3 5-6 8 11 13 16 18 21-22 24-26 28 30 32-35
fetal liver-spleen	Columbia University	FLS003	10 17-18 26
fetal liver	Invitrogen	FLV001	14 17 24 33 35
fetal liver	Clontech	FLV004	10 14 28
fetal muscle	Invitrogen	FMS001	3 6 10 26 31
fetal muscle	Invitrogen	FMS002	10 25 30
fetal skin	Invitrogen	FSK001	3 5-6 10 16-17 21-23 25-26 28 30-31
fetal skin	Invitrogen	FSK002	10 18 26
fetal spleen	BioChain	FSP001	10
umbilical cord	BioChain	FUC001	3 10 20 24 26 30 33
fetal brain	GIBCO	HFB001	3-6 8 10-11 15 18 21-24 35
infant brain	Columbia University	IB2002	2 4 7-8 10 14 16 18-19 21-23 26 28 35
infant brain	Columbia University	IB2003	3 5 8-10 16 20 23 26 29 35
infant brain	Columbia University	IBM002	19
infant brain	Columbia University	IBS001	26 35
lung, fibroblast	Strategene	LFB001	10 21-22 30 33
lung tumor	Invitrogen	LGT002	2 10-11 14-15 24 28 30 35
lymphocytes	ATCC	LPC001	1 3 11 18 28 35
leukocyte	GIBCO	LUC001	1-3 5-7 11 16 18 21-22 24-26 28 35
melanoma from cell line ATCC #CRL 1424	Clontech	MEL004	2 24 26
mammary gland	Invitrogen	MMG001	2 6 10 12 14 16 18-19 24 26 28 31 35
induced neuron cells	Strategene	NTD001	15 23
neuronal cells	Strategene	NTU001	10 26
placenta	Clontech	PLA003	33
prostate	Clontech	PRT001	7-8 10-11 14 16 21-22 24
rectum	Invitrogen	REC001	6 26 28
salivary gland	Clontech	SAL001	10 16 21-22 35
skin fibroblast	ATCC	SFB001	10
small intestine	Clontech	SIN001	3 5 10 14 24 30
skeletal muscle	Clontech	SKM001	10 15
spinal cord	Clontech	SPC001	18 26 28
adult spleen	Clontech	SPLc01	1 14 26
stomach	Clontech	STO001	16
thalamus	Clontech	THA002	2-3 24 26

TABLE 1

TISSUE ORIGIN	RNA SOURCE	HYSEQ LIBRARY NAME	SEQ ID NOS:
thymus	Clontech	THMc02	3 6-7 14 17 26 31-32 35
thyroid gland	Clontech	THR001	3 6 8 10-11 17-18 21-22 24 26 28 33
trachea	Clontech	TRC001	10 21-22 33
uterus	Clontech	UTR001	10-11 24

TABLE 2

SEQ ID NO:	CORRESPONDING SEQ ID NO. IN U.S.S.N 09/574,454	ACCESSION NUMBER	DESCRIPTION	SMITH-WATERMAN SCORE	% IDENTITY
1	1621	AL049785	Homo sapiens hypothetical protein	761	100
2	1724	U20657	Homo sapiens ubiquitin protease	719	47
3	2106	Y14494	Homo sapiens aralar1	3462	99
4	2500	U00051	Caenorhabditis elegans coded for by C. elegans cDNA yk50b2.5; coded for by C. elegans cDNA CEESV26F; similar to lipases over a short region	414	33
5	2501	X62575	Drosophila melanogaster ubiquitin-conjugating enzyme	78	26
6	2507	AB017005	Homo sapiens PMS2L14	1552	99
7	2520	AF222766	Bos taurus ankyrin 1	835	31
8	2555	AB039670	Homo sapiens ALEX1	2332	100
9	3766	U51000	Mus musculus DLX-1	1319	98
10	3935	X66405	Mus musculus collagen alpha1 type VI-precursor	5011	89
11	4297	AF208856	Homo sapiens BM-014	1012	99
12	4333	X62677	Oryctolagus cuniculus retrovirus related reverse transcriptase	75	52
13	4449	M14912	Homo sapiens pol	132	86
14	4562	AB024028	Arabidopsis thaliana uridine kinase-like protein	1021	45
15	4591	L11275	Saccharomyces cerevisiae selected as a weak suppressor of a mutant of the subunit AC40 of DNA dependant RNA polymerase I and III	97	19
16	4614	AL110500	Caenorhabditis elegans Y87G2A.13	260	24
17	4711	A14829	Homo sapiens preproapolipoprotein	974	100

TABLE 2

SEQ ID NO:	CORRESPONDING SEQ ID NO. IN U.S.S.N 09/574,454	ACCESSION NUMBER	DESCRIPTION	SMITH-WATERMAN SCORE	% IDENTITY
18	4726	AF151086	Homo sapiens HSPC252	1223	81
19	4737	AF083116	Homo sapiens paraneoplastic cancer-testis-brain antigen	122	42
20	4762	AC006963	Homo sapiens similar to Kelch proteins; similar to BAA77027 (PID:g4650844)	1226	43
21	4790	L39995	Ceratovacuna lanigera cytochrome oxidase I	76	28
22	4790	L39995	Ceratovacuna lanigera cytochrome oxidase I	76	28
23	4828	AB035266	Homo sapiens neurexin II	9034	100
24	4999	M37194	Rattus norvegicus clathrin-associated protein 17	704	91
25	5155	AF217516	Homo sapiens uncharacterized bone marrow protein BM040	669	100
26	5244	U35776	Rattus norvegicus ADP-ribosylation factor 1-directed GTPase activating protein	1849	84
27	5280	AF159852	Drosophila melanogaster RNA-binding protein Smaug	259	44
28	5502	Y13247	Homo sapiens FB19 protein	5157	99
29	5813	AF116638	Homo sapiens PRO1546	70	48
30	5868	L21013	Dictyostelium discoideum RabC	126	25
31	5890	AL031033	Homo sapiens C321D2.4 (novel protein)	960	100
32	6070	X58236	Homo sapiens 36/8-8 fusion protein with epitope for anti-lectin antibody	90	73
33	6245	AF095593	Homo sapiens caveolin-1	946	100
34	6277	AF133521	Libellula pulchella troponin T	45	43

TABLE 2

SEQ ID NO:	CORRESPONDING SEQ ID NO. IN U.S.S.N 09/574,454	ACCESSION NUMBER	DESCRIPTION	SMITH-WATERMAN SCORE	% IDENTITY
35	6298	AF161491	Homo sapiens HSPC142	1379	100

TABLE 3

SEQ ID NO:	ACCESSION NO.	DESCRIPTION	RESULTS*
1	DM01117	2 kw TRANSPOSASE WITHIN TRANSPOSITION VASOTOCIN.	DM01117A 11.17 9.173e-06 52-70
2	BL00972	Ubiquitin carboxyl-terminal hydrolases family 2 proteins.	BL00972A 11.93 5.091e-19 498-516 BL00972D 22.55 4.000e-15 1148-1173 BL00972C 16.48 9.143e-12 642-657 BL00972E 20.72 4.462e-11 1182-1204 BL00972B 9.45 4.176e-09 584-594
3	BL00215	Mitochondrial energy transfer proteins.	BL00215A 15.82 1.600e-15 333-358 BL00215A 15.82 1.794e-14 433-458 BL00215A 15.82 4.441e-14 525-550 BL00215B 10.44 3.250e-10 569-582 BL00215B 10.44 1.000e-08 381-394
4	PF00756	Putative esterase.	PF00756C 14.12 2.084e-09 206-236
5	PR00701	60KD INNER MEMBRANE PROTEIN SIGNATURE	PR00701I 8.59 6.256e-06 19-43
6	BL00058	DNA mismatch repair proteins mutL / hexB / PMS1 proteins.	BL00058A 20.73 5.125e-33 108-159 BL00058B 16.83 3.721e-29 172-211 BL00058C 18.05 7.568e-16 240-258 BL00058C 18.05 8.235e-13 36-54
7	PF00023	Ank repeat proteins.	PF00023A 16.03 7.600e-13 177-193 PF00023A 16.03 8.286e-11 144-160 PF00023A 16.03 2.500e-10 760-776 PF00023B 14.20 5.000e-10 339-349 PF00023B 14.20 5.500e-10 239-249 PF00023B 14.20 6.500e-10 471-481 PF00023B 14.20 8.000e-10 823-833 PF00023A 16.03 8.875e-10 554-570 PF00023A 16.03 1.000e-09 45-61 PF00023B 14.20 2.227e-09 438-448 PF00023B 14.20 2.636e-09 272-282 PF00023B 14.20 2.636e-09 583-593 PF00023A 16.03 5.821e-09 587-603 PF00023B 14.20 7.955e-09 405-415 PF00023A 16.03 9.679e-09 210-226
8	DM00892	3 RETROVIRAL PROTEINASE.	DM00892B 9.78 4.767e-06 27-33
9	PR00031	LAMBDA AND OTHER REPRESSOR HELIX-TURN-HELIX SIGNATURE	PR00031B 16.29 1.643e-17 166-183 PR00031A 8.77 9.400e-09 157-167
10	PR00159	2FE-2S FERREDOXIN SIGNATURE	PR00159B 8.50 7.882e-09 604-612
11	PR00138	MATRIXIN SIGNATURE	PR00138D 16.56 4.517e-09 77-103
12	BL00597	Plant lipid transfer proteins.	BL00597B 12.41 7.955e-06 11-40
13	BL00366	Uricase proteins.	BL00366A 11.55 9.780e-06 25-39
15	PF00602	Influenza RNA-dependant RNA polymerase subunit PB1.	PF00602E 11.62 5.160e-06 95-142
16	DM01688	2 POLY-IG RECEPTOR.	DM01688K 17.19 5.938e-06 321-360

* Results include in order: accession number subtype; raw score; p-value; position of signature in amino acid sequence.

TABLE 3

SEQ ID NO:	ACCESSION NO.	DESCRIPTION	RESULTS*
17	PD02807	APOLIPOPROTEIN E PRECURSOR APO-E GLYCOPROTEIN PLAS.	PD02807D 7.99 5.534e-09 105-155
18	DM00303	6 LEA 11-MER REPEAT REPEAT.	DM00303B 21.87 9.173e-07 217-252
19	BL00048	Protamine P1 proteins.	BL00048 6.39 8.763e-09 106-133 BL00048 6.39 9.663e-09 108-135
20	PF00651	BTB (also known as BR-C/Ttk) domain proteins.	PF00651 15.00 9.182e-15 76-89
21	PR00701	60KD INNER MEMBRANE PROTEIN SIGNATURE	PR00701B 15.26 7.188e-07 177-199
22	PR00701	60KD INNER MEMBRANE PROTEIN SIGNATURE	PR00701B 15.26 7.188e-07 177-199
23	DM00060	338 kw NEUREXIN ALPHA III CYSTEINE.	DM00060 6.92 3.925e-10 210-220
24	PR00317	EPENDYMIN SIGNATURE	PR00317F 10.90 5.935e-09 117-132
25	PR00305	14-3-3 PROTEIN ZETA SIGNATURE	PR00305F 15.95 7.150e-06 226-256
26	PR00405	HIV REV INTERACTING PROTEIN SIGNATURE	PR00405A 17.71 2.286e-20 19-39 PR00405B 11.83 6.077e-15 38-56 PR00405C 19.41 4.000e-13 59-81
28	PR00334	HMW KININOGEN SIGNATURE	PR00334B 8.69 9.914e-09 867-891
29	PR00551	2-S GLOBULIN FAMILY SIGNATURE	PR00551E 10.27 6.211e-06 12-27
30	PR00300	ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT SIGNATURE	PR00300A 9.56 7.896e-09 6-25
31	PR00467	MAMMALIAN LIPOXYGENASE SIGNATURE	PR00467E 9.00 1.000e-05 27-47
33	BL01210	Caveolins proteins.	BL01210B 13.92 1.000e-40 91-141 BL01210A 17.61 6.000e-39 43-82 BL01210C 17.86 8.579e-34 141-173
34	PR00658	CD44 ANTIGEN PRECURSOR SIGNATURE	PR00658D 8.52 9.919e-06 12-32
35	PR00512	5-HYDROXYTRYPTAMINE 1A RECEPTOR SIGNATURE	PR00512E 10.80 4.884e-06 30-47

TABLE 4

SEQ ID NO:	pFAM NAME	DESCRIPTION	p-value	pFAM SCORE
2	zf-MYND	MYND finger	1.1e-06	35.6
3	efhand	EF hand	0.0016	25.1
6	DNA_mis_repair	DNA mismatch repair protein	8.2e-36	132.4
7	ank	Ank repeat	3.6e-197	668.4
9	homeobox	Homeobox domain	1.5e-29	111.6
10	Collagen	Collagen triple helix repeat (20 copies)	7.5e-46	165.8
14	PRK	Phosphoribulokinase	9.5e-06	-24.1
17	Apolipoprotein	Apolipoprotein A1/A4/E family	5.2e-113	388.8
18	BTB	BTB/POZ domain	2.6e-25	97.5
20	Kelch	Kelch motif	1.7e-22	88.2
23	EGF	EGF-like domain	3.5e-09	43.9
24	Clat_adaptor_s	Clathrin adaptor complex small chain	5.7e-88	305.7
26	ArfGap	Putative GTP-ase activating protein for Arf	6.3e-62	219.2
27	SAM	SAM domain (Sterile alpha motif)	0.003	20.7
30	ras	Ras family	0.0058	-103.2
33	Caveolin	Caveolin	3.7e-120	412.6

TABLE 5

SEQ ID NO:	POSITION OF SIGNAL IN AMINO ACID SEQUENCE	maxS (MAXIMUM SCORE)	meanS (MEAN SCORE)
10	1-19	0.968	0.899
12	1-23	0.960	0.861
16	1-28	0.983	0.763
17	1-18	0.966	0.921
23	1-28	0.963	0.881
29	1-20	0.940	0.677

CLAIMS

WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 – 35, a mature protein coding portion of SEQ ID NO: 1 – 35, an active domain of SEQ ID NO: 1 – 35, and complementary sequences thereof.

2. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide hybridizes to the polynucleotide of claim 1 under stringent hybridization conditions.

3. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide has greater than about 90% sequence identity with the polynucleotide of claim 1.

4. The polynucleotide of claim 1 wherein said polynucleotide is DNA.

5. An isolated polynucleotide of claim 1 wherein said polynucleotide comprises the complementary sequences.

6. A vector comprising the polynucleotide of claim 1.

7. An expression vector comprising the polynucleotide of claim 1.

8. A host cell genetically engineered to comprise the polynucleotide of claim 1.

9. A host cell genetically engineered to comprise the polynucleotide of claim 1 operatively associated with a regulatory sequence that modulates expression of the polynucleotide in the host cell.

10. An isolated polypeptide, wherein the polypeptide is selected from the group consisting of:

(a) a polypeptide encoded by any one of the polynucleotides of claim 1; and

(b) a polypeptide encoded by a polynucleotide hybridizing under stringent conditions with any one of SEQ ID NO: 1 – 35.

11. A composition comprising the polypeptide of claim 10 and a carrier.

12. An antibody directed against the polypeptide of claim 10.

13. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

a) contacting the sample with a compound that binds to and forms a complex with the polynucleotide of claim 1 for a period sufficient to form the complex; and

b) detecting the complex, so that if a complex is detected, the polynucleotide of claim 1 is detected.

14. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

a) contacting the sample under stringent hybridization conditions with nucleic acid primers that anneal to the polynucleotide of claim 1 under such conditions;

b) amplifying a product comprising at least a portion of the polynucleotide of claim 1; and

c) detecting said product and thereby the polynucleotide of claim 1 in the sample.

15. The method of claim 14, wherein the polynucleotide is an RNA molecule and the method further comprises reverse transcribing an annealed RNA molecule into a cDNA polynucleotide.

16. A method for detecting the polypeptide of claim 10 in a sample, comprising:

a) contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex; and

b) detecting formation of the complex, so that if a complex formation is detected, the polypeptide of claim 10 is detected.

17. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

a) contacting the compound with the polypeptide of claim 10 under conditions sufficient to form a polypeptide/compound complex; and

b) detecting the complex, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

18. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

a) contacting the compound with the polypeptide of claim 10, in a cell, under conditions sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and

b) detecting the complex by detecting reporter gene sequence expression, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

19. A method of producing the polypeptide of claim 10, comprising,

a) culturing a host cell comprising a polynucleotide sequence selected from the group consisting of a polynucleotide sequence of SEQ ID NO: 1-35, a mature protein coding portion of SEQ ID NO: 1-35, an active domain of SEQ ID NO: 1-35, complementary sequences thereof and a polynucleotide sequence hybridizing under stringent conditions to SEQ ID NO: 1-35, under conditions sufficient to express the polypeptide in said cell; and

b) isolating the polypeptide from the cell culture or cells of step (a).

20. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of any one of the polypeptides from the Sequence Listing, the mature protein portion thereof, or the active domain thereof.

21. The polypeptide of claim 20 wherein the polypeptide is provided on a polypeptide array.

22. A collection of polynucleotides, wherein the collection comprising the sequence information of at least one of SEQ ID NO: 1 – 35.

23. The collection of claim 22, wherein the collection is provided on a nucleic acid array.

24. The collection of claim 23, wherein the array detects full-matches to any one of the polynucleotides in the collection.

25. The collection of claim 23, wherein the array detects mismatches to any one of the polynucleotides in the collection.

26. The collection of claim 22, wherein the collection is provided in a computer-readable format.

27. A method of treatment comprising administering to a mammalian subject in need thereof a therapeutic amount of a composition comprising a polypeptide of claim 10 or 20 and a pharmaceutically acceptable carrier.

28. A method of treatment comprising administering to a mammalian subject in need thereof a therapeutic amount of a composition comprising an antibody that specifically binds to a polypeptide of claim 10 or 20 and a pharmaceutically acceptable carrier.

ABSTRACT OF THE INVENTION

The present invention provides novel nucleic acids, novel polypeptide sequences encoded by these nucleic acids and uses thereof.

5

Express Mail No.: EK916750871US
Docket No.: 789CIP2C

DECLARATION FOR PATENT APPLICATION AND POWER OF ATTORNEY

As [a] below named inventor(s), I/we hereby declare that:

Y. Tom Tang, Chenghua Liu, Ping Zhou, Vinod Asundi, Feiyan Ren, Qing A. Zhao, Jie Zhang, Jian-Rui Wang, Tom Wehrman, Radoje T. Drmanac

My/our residence, post office address and citizenship is/are as stated below next to my/our name(s).

I/we believe I/we am/are an/the original, first and sole/joint inventor of the subject matter which is claimed and for which a patent is sought on the invention entitled: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES, the specification of which

__X__ is attached hereto.

____ was filed on [date] as Application Serial Number [ ] and was amended on [date].

I/We hereby state that I/we have reviewed and understand the contents of the above-identified specification, including the claims as amended by any amendment referred to above.

I/We acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, Section 1.56(a).

I/We hereby claim foreign priority benefits under Title 35, United States Code, § 119 of any foreign application(s) for patent or inventor's certificate, listed below and so identified, and I/we have also identified below any foreign application for patent or inventor's certificate on this invention filed by me or my legal representatives or assigns and having a filing date before that of the application on which priority is claimed:

NUMBER	COUNTRY	DAY/MONTH/ YEAR FILED	PRIORITY CLAIMED - YES OR NO

I/We hereby claim the benefit under Title 35, United States Code, § 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, § 112, I/we acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, § 1.56(a) which occurred between the filing date of the prior application and the national or PCT international filing date of this application.

SERIAL NUMBER	FILING DATE	STATUS
09/574,454	May 19, 2000	Pending
09/519,705	March 07, 2000	Pending

I/We hereby declare that all statements made herein of my/our own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under § 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

I/We hereby appoint the following attorneys and agents, with full power of substitution and revocation, to prosecute this application and to transact all business in the United States Patent and Trademark Office connected therewith and request that all correspondence and telephone calls with respect to this application be directed to Leslie A. Mooi, HYSEQ, INC., 670 Almanor Avenue, Sunnyvale, CA 94085, Telephone No. (408) 524-8100:

ATTORNEY	REGISTRATION NO.
Petrina S. Hsi	38,496
Leslie A. Mooi	37,047

Full name of first joint inventor: Y. Tom Tang

Inventor's signature: _____ Date: _____

Residence and Post Office Address: 4230 Ranwick Court, San Jose, CA 95118

Citizenship: United States of America

Full name of second joint inventor: Chenghua Liu

Inventor's signature: _____ Date: _____

Residence and Post Office Address: 1125 Ranchero Way, Apt. # 14, San Jose, CA 95117

Citizenship: People's Republic of China

Full name of third joint inventor: Ping Zhou

Inventor's signature: _____ Date: _____

Residence and Post Office Address: 1461 Japaur Lane, San Jose, CA 95132

Citizenship: People's Republic of China

Full name of fourth joint inventor: Vinod Asundi

Inventor's signature: _____ Date: _____

Residence and Post Office Address: 709 Foster City Blvd., Foster City, CA 94404

Citizenship: United States of America

Full name of fifthjoint inventor: Feiyan Ren

Inventor's signature: _____ Date: _____

Residence and Post Office Address: 20685 Garden Manor Court, Cupertino, CA 95014

Citizenship: People's Republic of China

Full name of sixth joint inventor: Qing A. Zhao

Inventor's signature: _____ Date: _____

Residence and Post Office Address: 1028 S. de Anza Blvd., Apt. B-210, San Jose, CA 95129

Citizenship: People's Republic of China

Full name of
seventh joint inventor: Jie Zhang

Inventor's signature: _____ Date: _____

Residence and Post
Office Address: 20800 Homestead Road, #38B, Cupertino, CA 95014

Citizenship: People's Republic of China

Full name of
eighth joint inventor: Jian-Rui Wang

Inventor's signature: _____ Date: _____

Residence and Post
Office Address: 744 Stendhal Lane, Cupertino, CA 95014

Citizenship: People's Republic of China

Full name of
ninth joint inventor: Tom Wehrman

Inventor's signature: _____ Date: _____

Residence and Post
Office Address: 300 Pasteur Drive, Edwards R314, Stanford University Medical Center, Stanford, CA 94305

Citizenship: United States of America

Full name of
tenth joint inventor: Radoje T. Drmanac

Inventor's signature: _____ Date: _____

Residence and Post
Office Address: 850 East Greenwich Place, Palo Alto, CA 94303

Citizenship: Yugoslavia

SEQUENCE LISTING

<110> Tang, Y. Tom
Liu, Chenghua
Zhou, Ping
Asundi, Vinod
Ren, Feiyan
Zhao, Qing A.
Zhang, Jie
Wang, Jian-Rui
Wehrman, Tom
Drmanac, Radoje T.

<120> Novel Nucleic Acids and
Polypeptides

<130> 789CIP2C

<140> To Be Assigned
<141>

<150> 09/574,454
<151> 2000-05-19

<150> 09/519,705
<151> 2000-03-07

<160> 35

<170> pt_FL_genes Version 2.0

<210> 1
<211> 2235
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (258)..(671)

<400> 1
taagcttgcg gccgccccgg ccgcccccgc gggggacacc tcctcgccgc cacagcttta 60

ggaaacacct ctacctcctg cgaggcctcc cgggctccgg gaaaactaca ctggccagac 120

aattgcagca tgactttccc agggccctga ttttcagcac ggatgatttt ttcttcaggg 180

```
aagatggtgc ctatgagttc aatcctgact tcctggagga agctcatgaa tggaaccaaa     240

aaagagcaag aaaagca   atg agg aat ggc ata tcc ccc att att att gat     290
                     Met Arg Asn Gly Ile Ser Pro Ile Ile Ile Asp
                       1               5                  10

aat acc aac ctc cac gcc tgg gaa atg aag ccc tat gca gtc atg gca      338
Asn Thr Asn Leu His Ala Trp Glu Met Lys Pro Tyr Ala Val Met Ala
            15                  20                  25

ctt gaa aat aac tat gaa gtt ata ttc cga gaa cct gac act cgc tgg      386
Leu Glu Asn Asn Tyr Glu Val Ile Phe Arg Glu Pro Asp Thr Arg Trp
        30                  35                  40

aaa ttc aac gtt caa gag tta gca aga aga aac att cat ggt gtc tca      434
Lys Phe Asn Val Gln Glu Leu Ala Arg Arg Asn Ile His Gly Val Ser
    45                  50                  55

aga gaa aaa atc cac cga atg aaa gaa cgg tat gaa cac gat gtt act      482
Arg Glu Lys Ile His Arg Met Lys Glu Arg Tyr Glu His Asp Val Thr
60                  65                  70                  75

ttt cac agt gtg ctt cat gca gaa aag cca agc aga atg aac aga aac      530
Phe His Ser Val Leu His Ala Glu Lys Pro Ser Arg Met Asn Arg Asn
                80                  85                  90

cag gac agg aat aat gca ttg cct tcc aac aat gcc aga tac tgg aat      578
Gln Asp Arg Asn Asn Ala Leu Pro Ser Asn Asn Ala Arg Tyr Trp Asn
            95                  100                 105

tcc tac aca gag ttt cca aac cgg agg gcc cac ggt gga ttt aca aat      626
Ser Tyr Thr Glu Phe Pro Asn Arg Arg Ala His Gly Gly Phe Thr Asn
        110                 115                 120

gag agc tcc tat cac aga agg ggc ggt tgt cac cat gga tat tag agg      674
Glu Ser Ser Tyr His Arg Arg Gly Gly Cys His His Gly Tyr  *
    125                 130                 135

cctatcttac agccaggcag aattttccta agtcagtttc tacttcagtt tttgttattt    734

tttgttgcat tttagtcaga gctccaattc cagtgtaaat agctgaactc aaaagtttct    794

gagcaaagtc attatattca ctttcttcac caaaatttgt taaagtgctt ctatatgcat    854

ggtctgatgc tgggaattct gcagatttga gtaaacagtc tctttctcta gggtaagaat    914

ttgaaaccaa aacttgagaa cacacccaag aatatattta cataggttca tagatgaaat    974

aaagtgttta tattatatat aagcttcagt accatttgct ctgaagtgat ctatttattt   1034

tttcaggaaa ttcatctcca tcggtaaagt tgggaaggtg gagagaagtg gtgggggggc   1094

attgctactt atcaaagtgc cattgctact ttgataatct atgtatctaa aaatgtgaga   1154

tgtgcgactc ttatgatact gattttcctt taatgttaat atgccagaaa gcatacatct   1214
```

aagggaacat tgtccttcaa agtagacact ttgggaagtt atttctttat tttaatgatg 1274

tatcattgtt aaaaatgctg tcaaatcctt aatagctaca ggagctactg agggaaatca 1334

gtgtcattat ttaaagtcac gccttgtgtt tttactactt tattcagcag gattaaacct 1394

gaataacttt tggctgttgt gctaatagtg taaataaaat aagcctgcct tcataaaaca 1454

ctaactttta aaaggaataa acgacttcta aaattatgcc tattaacatg tgtaattagt 1514

cggcagctca aatgtttggg agtgcaagaa attaggcacc ccaggatata ggtcatacag 1574

ggatatataa aagccatgct cattacaaaa tgagcagttg atgttttatg tggcattaag 1634

acaatcaagt cctcacaact ctggaatgtc ttcttatact gatgctgaat ttatgaatcc 1694

aaattaattt ccaacaggtt ggaatcagat ttaatgtgag atcatgatag acaagaccac 1754

agaggacgta tgctctattt cttgttggcc aacagcttct ttctaatgtt ctgtgaaaaa 1814

ttattttaag tgtcttatat aatggtgctt ttatggttat taaaaattgt aaatggtatc 1874

acatttatat ggatttgtca ttggatcttt ttttggttca acaataaaaa aatttaatta 1934

cctaaatgcc aagaaactca acaatatacc agtttttctg tatcacaggc ttatttacca 1994

gtctttttt taataaatag gaatcgtaaa ggtaatgaca aaagcagcct tataatttag 2054

ttgcttatat atttgatctg tgtacatgag actgttttaa cgttatctga cactactgaa 2114

acctgctcga catctccatg actaccaaca ccatgtgtaa tgttttcttc actaacattt 2174

taaaaactgg tatctccttt gagtaagttt ggctgacaat agtaaatccc aatgaatcta 2234

a 2235

<210> 2
<211> 4600
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (349)..(4305)

<400> 2
ttcaaggtca cgggccgtgc cagtccctac ccctagtgcc ccgcagcgtg ccagcccact 60

gatgccggca ggccgttacg ctatgcctgt ggcagcctcc cacctgggag aggactggcg 120

gccacgaggg caacttcgtg ggagaggtgg tggcgggctc ggggttcaga ggctgaagct 180

gggtcgcagc tcgaccgggg cgctgatctg tggctgtccg ttccttgctg gagaatttgg 240

```
ccacaaagag ctgccaagat agctgggcca ggaagaaagc gccgcagccc tgacccagac    300

gctgttgccg accccggggc actctggctg tcgaccaagc ggctcaag atg tct ggc     357
                                                      Met Ser Gly
                                                        1

ggg gcc agt gcc aca ggc cca agg aga ggg ccc cca gga ctg gag gac       405
Gly Ala Ser Ala Thr Gly Pro Arg Arg Gly Pro Pro Gly Leu Glu Asp
      5                  10                  15

acc act agt aag aag aag cag aag gat cga gca aac cag gag agc aag       453
Thr Thr Ser Lys Lys Lys Gln Lys Asp Arg Ala Asn Gln Glu Ser Lys
 20                  25                  30                  35

gat gga gat cct agg aaa gag aca ggg tct cga tat gtt gcc cag gct       501
Asp Gly Asp Pro Arg Lys Glu Thr Gly Ser Arg Tyr Val Ala Gln Ala
                 40                  45                  50

ggt ctt gaa cct ctg gcc tca ggt gat cct tct gcc tca gcc tcc cat       549
Gly Leu Glu Pro Leu Ala Ser Gly Asp Pro Ser Ala Ser Ala Ser His
             55                  60                  65

gca gct ggg atc aca ggc tca cgc cac cgt acc cgg ctg ttc ttt cct       597
Ala Ala Gly Ile Thr Gly Ser Arg His Arg Thr Arg Leu Phe Phe Pro
         70                  75                  80

tca tcg tca ggg tca gca tcc act cct caa gag gag cag acc aaa gag       645
Ser Ser Ser Gly Ser Ala Ser Thr Pro Gln Glu Glu Gln Thr Lys Glu
     85                  90                  95

gga gct tgt gaa gac cct cat gat ctc ttg gct act ccc act cca gag       693
Gly Ala Cys Glu Asp Pro His Asp Leu Leu Ala Thr Pro Thr Pro Glu
100                 105                 110                 115

ttg ttg ctc gat tgg agg cag agt gca gaa gag gtg att gtc aag ctt       741
Leu Leu Leu Asp Trp Arg Gln Ser Ala Glu Glu Val Ile Val Lys Leu
                120                 125                 130

cgt gtg gga gta ggt ccc ctg cag ctg gag gat gta gat gct gct ttc       789
Arg Val Gly Val Gly Pro Leu Gln Leu Glu Asp Val Asp Ala Ala Phe
            135                 140                 145

aca gat aca gac tgt gtg gtg cgg ttt gca ggt ggt cag cag tgg ggt       837
Thr Asp Thr Asp Cys Val Val Arg Phe Ala Gly Gly Gln Gln Trp Gly
        150                 155                 160

ggt gtc ttc tat gct gag ata aaa agc tct tgt gct aaa gtg caa acc       885
Gly Val Phe Tyr Ala Glu Ile Lys Ser Ser Cys Ala Lys Val Gln Thr
    165                 170                 175

cgc aag ggc agt ctc ctg cac ctg aca ctg ccc aaa aag gtg cct atg       933
Arg Lys Gly Ser Leu Leu His Leu Thr Leu Pro Lys Lys Val Pro Met
180                 185                 190                 195

ctc acg tgg ccc tcc ctc ctg gtt gag gct gat gaa cag ctt tgc ata       981
Leu Thr Trp Pro Ser Leu Leu Val Glu Ala Asp Glu Gln Leu Cys Ile
                200                 205                 210
```

```
cca ccg ctg aac tcc caa acc tgc ctc ctg ggc tca gag gag aat tta     1029
Pro Pro Leu Asn Ser Gln Thr Cys Leu Leu Gly Ser Glu Glu Asn Leu
            215                 220                 225

gcc cct ttg gca gga gag aaa gca gtg cct ccc ggg aat gac cca gtc     1077
Ala Pro Leu Ala Gly Glu Lys Ala Val Pro Pro Gly Asn Asp Pro Val
        230                 235                 240

tct cca gcc atg gtc cgg agc aga aac cct ggg aaa gat gac tgt gcc     1125
Ser Pro Ala Met Val Arg Ser Arg Asn Pro Gly Lys Asp Asp Cys Ala
    245                 250                 255

aag gag gag atg gca gtg gca gca gat gct gca acc ttg gtg gat gag     1173
Lys Glu Glu Met Ala Val Ala Ala Asp Ala Ala Thr Leu Val Asp Glu
260                 265                 270                 275

ccc gag tcg atg gtg aac ctg gcg ttt gtc aag aat gac tcg tat gag     1221
Pro Glu Ser Met Val Asn Leu Ala Phe Val Lys Asn Asp Ser Tyr Glu
                280                 285                 290

aag ggc ccg gat tca gtg gtg gtg cac gtg tac gtg aag gag atc tgc     1269
Lys Gly Pro Asp Ser Val Val Val His Val Tyr Val Lys Glu Ile Cys
            295                 300                 305

agg gac acc tca aga gta ctt ttc cgt gag cag gac ttc acg ctc atc     1317
Arg Asp Thr Ser Arg Val Leu Phe Arg Glu Gln Asp Phe Thr Leu Ile
        310                 315                 320

ttc cag acc agg gat gga aac ttc ctg agg ctg cac ccg ggc tgt ggg     1365
Phe Gln Thr Arg Asp Gly Asn Phe Leu Arg Leu His Pro Gly Cys Gly
    325                 330                 335

ccc cac acc acc ttc cgt tgg cag gtg aag ctc agg aat ctg att gag     1413
Pro His Thr Thr Phe Arg Trp Gln Val Lys Leu Arg Asn Leu Ile Glu
340                 345                 350                 355

cca gag cag tgc acc ttc tgt ttc acg gct tct cgc atc gac atc tgc     1461
Pro Glu Gln Cys Thr Phe Cys Phe Thr Ala Ser Arg Ile Asp Ile Cys
                360                 365                 370

ctt cgt aag agg cag agt cag cgc tgg ggg ggc ctg gag gcc ccg gct     1509
Leu Arg Lys Arg Gln Ser Gln Arg Trp Gly Gly Leu Glu Ala Pro Ala
            375                 380                 385

gca cga gtg ggt ggt gca aag gtt gcc gtg ccg aca ggt cca acc cct     1557
Ala Arg Val Gly Gly Ala Lys Val Ala Val Pro Thr Gly Pro Thr Pro
        390                 395                 400

ctg gat tca acc cca cca gga ggt gct ccc cac ccc ctg aca ggc cag     1605
Leu Asp Ser Thr Pro Pro Gly Gly Ala Pro His Pro Leu Thr Gly Gln
    405                 410                 415

gag gag gcc cgg gct gtg gag aag gat aaa tcc aag gca cga tct gag     1653
Glu Glu Ala Arg Ala Val Glu Lys Asp Lys Ser Lys Ala Arg Ser Glu
420                 425                 430                 435
```

```
gac aca ggg cta gac agt gtg gca acc cgc aca ccc atg gag cat gta     1701
Asp Thr Gly Leu Asp Ser Val Ala Thr Arg Thr Pro Met Glu His Val
                440                 445                 450

acc cca aag cca gag aca cac ctg gcc tcg ccc aag cct aca tgc atg     1749
Thr Pro Lys Pro Glu Thr His Leu Ala Ser Pro Lys Pro Thr Cys Met
            455                 460                 465

gtg cct ccc atg ccc cac agc cca gtt agt gga gac agc gtg gag gag     1797
Val Pro Pro Met Pro His Ser Pro Val Ser Gly Asp Ser Val Glu Glu
        470                 475                 480

gag gaa gag gaa gag aag aag gtg tgt ctg cca ggc ttc act ggc ctt     1845
Glu Glu Glu Glu Glu Lys Lys Val Cys Leu Pro Gly Phe Thr Gly Leu
    485                 490                 495

gtc aat tta ggc aac acc tgc ttc atg aac agc gtc att cag tct ctg     1893
Val Asn Leu Gly Asn Thr Cys Phe Met Asn Ser Val Ile Gln Ser Leu
500                 505                 510                 515

tcc aac act cgg gaa ctc cgg gac ttc ttc cat gac cgc tcc ttt gag     1941
Ser Asn Thr Arg Glu Leu Arg Asp Phe Phe His Asp Arg Ser Phe Glu
                520                 525                 530

gct gag atc aac tac aac aac cca cta ggg act ggt ggg cgt ctg gcc     1989
Ala Glu Ile Asn Tyr Asn Asn Pro Leu Gly Thr Gly Gly Arg Leu Ala
            535                 540                 545

att ggc ttt gcc gtg ctg ctt cgg gcg ctg tgg aag ggc acc cac cat     2037
Ile Gly Phe Ala Val Leu Leu Arg Ala Leu Trp Lys Gly Thr His His
        550                 555                 560

gcc ttc cag cct tcc aag ttg aag gcc att gtg gcg agt aag gcc agc     2085
Ala Phe Gln Pro Ser Lys Leu Lys Ala Ile Val Ala Ser Lys Ala Ser
    565                 570                 575

cag ttc aca ggc tat gca cag cat gat gcc cag gag ttc atg gct ttc     2133
Gln Phe Thr Gly Tyr Ala Gln His Asp Ala Gln Glu Phe Met Ala Phe
580                 585                 590                 595

ctg ctg gat ggg ctg cac gag gac ctg aat cgc att cag aac aag ccc     2181
Leu Leu Asp Gly Leu His Glu Asp Leu Asn Arg Ile Gln Asn Lys Pro
                600                 605                 610

tac aca gag acc gtg gat tca gat ggg cgg ccc gat gag gtg gta gct     2229
Tyr Thr Glu Thr Val Asp Ser Asp Gly Arg Pro Asp Glu Val Val Ala
            615                 620                 625

gag gaa gca tgg cag cgg cac aag atg agg aat gac tct ttc atc gtg     2277
Glu Glu Ala Trp Gln Arg His Lys Met Arg Asn Asp Ser Phe Ile Val
        630                 635                 640

gac cta ttt cag ggg cag tac aag tcg aag ctg gtg tgc cct gtg tgt     2325
Asp Leu Phe Gln Gly Gln Tyr Lys Ser Lys Leu Val Cys Pro Val Cys
    645                 650                 655

gcc aag gtc tcc atc act ttt gac ccg ttt ctt tat ctg ccg gtg ccc     2373
```

```
Ala Lys Val Ser Ile Thr Phe Asp Pro Phe Leu Tyr Leu Pro Val Pro
660                 665                 670                 675

ttg cca caa aag caa aag gtt ctc cct gtc ttt tat ttt gcc cga gag     2421
Leu Pro Gln Lys Gln Lys Val Leu Pro Val Phe Tyr Phe Ala Arg Glu
                680                 685                 690

ccc cac agc aag ccc atc aag ttc ctg gtg agc gtc agc aag gag aac     2469
Pro His Ser Lys Pro Ile Lys Phe Leu Val Ser Val Ser Lys Glu Asn
            695                 700                 705

tcc act gcg agc gaa gta ttg gac tcc ctc tct cag agt gtt cat gtg     2517
Ser Thr Ala Ser Glu Val Leu Asp Ser Leu Ser Gln Ser Val His Val
        710                 715                 720

aag cct gag aac ctg cgt ttg gcg gag gta att aag aat cgt ttt cat     2565
Lys Pro Glu Asn Leu Arg Leu Ala Glu Val Ile Lys Asn Arg Phe His
    725                 730                 735

cgt gtg ttc cta ccc tcc cac tca ctg gac act gtg tcc cca tct gat     2613
Arg Val Phe Leu Pro Ser His Ser Leu Asp Thr Val Ser Pro Ser Asp
740                 745                 750                 755

acg ctc ctc tgc ttt gag ctg cta tcc tca gag ttg gct aag gag cgg     2661
Thr Leu Leu Cys Phe Glu Leu Leu Ser Ser Glu Leu Ala Lys Glu Arg
                760                 765                 770

gta gtg gtg cta gag gtg caa cag cgc ccc cag gtg ccc agc gtc ccc     2709
Val Val Val Leu Glu Val Gln Gln Arg Pro Gln Val Pro Ser Val Pro
            775                 780                 785

atc tcc aag tgt gca gcc tgc cag cgg aag caa cag tcg gag gat gaa     2757
Ile Ser Lys Cys Ala Ala Cys Gln Arg Lys Gln Gln Ser Glu Asp Glu
        790                 795                 800

aag ctg aag cgc tgt acc cgg tgc tac cgt gtg ggc tac tgc aac cag     2805
Lys Leu Lys Arg Cys Thr Arg Cys Tyr Arg Val Gly Tyr Cys Asn Gln
    805                 810                 815

ctc tgc cag aaa acc cac tgg cct gac cac aag ggc ctc tgc cga cct     2853
Leu Cys Gln Lys Thr His Trp Pro Asp His Lys Gly Leu Cys Arg Pro
820                 825                 830                 835

gag aac att ggc tac ccc ttc ctg gtc agt gta cct gcc tca cgc ctc     2901
Glu Asn Ile Gly Tyr Pro Phe Leu Val Ser Val Pro Ala Ser Arg Leu
                840                 845                 850

act tat gcc cgc ctc gct cag ttg cta gag ggc tat gcc cgg tac tct     2949
Thr Tyr Ala Arg Leu Ala Gln Leu Leu Glu Gly Tyr Ala Arg Tyr Ser
            855                 860                 865

gtg agt gta ttc cag cca ccc ttt cag cca ggc cgc atg gcc ttg gag     2997
Val Ser Val Phe Gln Pro Pro Phe Gln Pro Gly Arg Met Ala Leu Glu
        870                 875                 880

tct cag agc cct ggc tgc acc aca ctg ctc tcc aca ggt tcc ctg gag     3045
Ser Gln Ser Pro Gly Cys Thr Thr Leu Leu Ser Thr Gly Ser Leu Glu
```

```
    885                 890                 895

gct ggg gac agc gag aga gac ccc att cag cca cct gag ctc cag ctg     3093
Ala Gly Asp Ser Glu Arg Asp Pro Ile Gln Pro Pro Glu Leu Gln Leu
900                 905                 910                 915

gtg acc cct atg gct gag ggg gac aca ggg ctt ccc cgg gtg tgg gca     3141
Val Thr Pro Met Ala Glu Gly Asp Thr Gly Leu Pro Arg Val Trp Ala
                920                 925                 930

gcc cct gac cgg ggt cct gtg ccc agc acc agt gga att tct tct gag     3189
Ala Pro Asp Arg Gly Pro Val Pro Ser Thr Ser Gly Ile Ser Ser Glu
            935                 940                 945

atg ctg gcc agt ggg ccc att gag gtt ggc tcc ttg cca gct ggc gag     3237
Met Leu Ala Ser Gly Pro Ile Glu Val Gly Ser Leu Pro Ala Gly Glu
        950                 955                 960

agg gtg tcc cga ccc gaa gct gct gtg cct ggg tac cag cat cca agt     3285
Arg Val Ser Arg Pro Glu Ala Ala Val Pro Gly Tyr Gln His Pro Ser
    965                 970                 975

gaa gct atg aat gcc cac aca ccc cag ttc ttc atc tat aaa att gat     3333
Glu Ala Met Asn Ala His Thr Pro Gln Phe Phe Ile Tyr Lys Ile Asp
980                 985                 990                 995

tca tcc aac cga gag cag cgg cta gag gac aaa gga gac acc cca ctg     3381
Ser Ser Asn Arg Glu Gln Arg Leu Glu Asp Lys Gly Asp Thr Pro Leu
                1000                1005                1010

gag ctg ggt gac gac tgt agc ctg gct ctc gtc tgg cgg aac aat gag     3429
Glu Leu Gly Asp Asp Cys Ser Leu Ala Leu Val Trp Arg Asn Asn Glu
            1015                1020                1025

cgc ttg cag gag ttt gtg ttg gta gcc tcc aag gag ctg gaa tgt gct     3477
Arg Leu Gln Glu Phe Val Leu Val Ala Ser Lys Glu Leu Glu Cys Ala
        1030                1035                1040

gag gat cca ggc tct gcc ggt gag gct gcc cgg gcc ggc cac ttc acc     3525
Glu Asp Pro Gly Ser Ala Gly Glu Ala Ala Arg Ala Gly His Phe Thr
    1045                1050                1055

ctg gac cag tgc ctc aac ctc ttc aca cgg cct gag gtg ctg gca ccc     3573
Leu Asp Gln Cys Leu Asn Leu Phe Thr Arg Pro Glu Val Leu Ala Pro
1060                1065                1070                1075

gag gag gcc tgg tac tgc cca cag tgc aaa cag cac cgt gag gcc tcc     3621
Glu Glu Ala Trp Tyr Cys Pro Gln Cys Lys Gln His Arg Glu Ala Ser
                1080                1085                1090

aag cag ctg ttg cta tgg cgc ctg cca aat gtt ctc atc gtg cag ctc     3669
Lys Gln Leu Leu Leu Trp Arg Leu Pro Asn Val Leu Ile Val Gln Leu
            1095                1100                1105

aag cgc ttc tcc ttt cgt agt ttt atc tgg cgt gac aag atc aat gac     3717
Lys Arg Phe Ser Phe Arg Ser Phe Ile Trp Arg Asp Lys Ile Asn Asp
        1110                1115                1120
```

```
ttg gtg gag ttc cct gtt agg aac ctg gac ctg agc aag ttc tgc att     3765
Leu Val Glu Phe Pro Val Arg Asn Leu Asp Leu Ser Lys Phe Cys Ile
   1125                1130                1135

ggt cag aaa gag gag cag ctg ccc agc tac gat cta tat gct gtc atc     3813
Gly Gln Lys Glu Glu Gln Leu Pro Ser Tyr Asp Leu Tyr Ala Val Ile
1140                1145                1150                1155

aac cac tat gga ggc atg att ggt ggc cac tac act gcc tgt gca cgc     3861
Asn His Tyr Gly Gly Met Ile Gly Gly His Tyr Thr Ala Cys Ala Arg
               1160                1165                1170

ctg ccc aat gat cgt agc agt cag cgc agt gac gtg ggc tgg cgc ttg     3909
Leu Pro Asn Asp Arg Ser Ser Gln Arg Ser Asp Val Gly Trp Arg Leu
           1175                1180                1185

ttt gat gac agc aca gtg aca acg gta gac gag agc cag gtt gtg acg     3957
Phe Asp Asp Ser Thr Val Thr Thr Val Asp Glu Ser Gln Val Val Thr
       1190                1195                1200

cgt tat gcc tat gta ctc ttc tac cgc cgg cgg aac tct cct gtg gag     4005
Arg Tyr Ala Tyr Val Leu Phe Tyr Arg Arg Arg Asn Ser Pro Val Glu
   1205                1210                1215

agg ccc ccc agg gca ggt cac tct gag cac cac cca gac cta ggc cct     4053
Arg Pro Pro Arg Ala Gly His Ser Glu His His Pro Asp Leu Gly Pro
1220                1225                1230                1235

gca gct gag gct gct gcc agc cag gct tcc cgg att tgg cag gag ctg     4101
Ala Ala Glu Ala Ala Ala Ser Gln Ala Ser Arg Ile Trp Gln Glu Leu
               1240                1245                1250

gag gct gag gag gag ccg gtg cct gag ggg tct ggg ccc ctg ggt ccc     4149
Glu Ala Glu Glu Glu Pro Val Pro Glu Gly Ser Gly Pro Leu Gly Pro
           1255                1260                1265

tgg ggg ccc caa gac tgg gtg ggc ccc cta cca cgt ggc cct acc aca     4197
Trp Gly Pro Gln Asp Trp Val Gly Pro Leu Pro Arg Gly Pro Thr Thr
       1270                1275                1280

cca gat gag ggc tgc ctc cgg tac ttt gtc ctg ggc acc gtg gcg gct     4245
Pro Asp Glu Gly Cys Leu Arg Tyr Phe Val Leu Gly Thr Val Ala Ala
   1285                1290                1295

ttg gtg gcc ctc gtg ctc aac gtg ttc tat cct ctg gta tcc cag agt     4293
Leu Val Ala Leu Val Leu Asn Val Phe Tyr Pro Leu Val Ser Gln Ser
1300                1305                1310                1315

cgc tgg aga tga gct cgcctgcagg cagctgctgt gagctggcct acctgcctgc     4348
Arg Trp Arg  *

cccaggccat gcctgccttt gttgtgggga acacctctgg gctttgggcc tcagcttatg   4408

catctggtgg gagagggtgg ggaggttgtg gcccctgcag gggcagagta tcctagggtg   4468
```

```
tgtatccatc tggctgtctg tccattcatc ctgctgctct gacccttggc ctcaggcttg   4528

gccctgccca agctacttcc tgtacttaaa agtgttaata aaaccagact attcaggccc   4588

aaaaaaaaaa aa                                                       4600
```

<210> 3
<211> 2663
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (115)..(2157)

<400> 3

```
cggaattccc gggtcgacga tttcgtcgcg gccggctgag tcctctccag ccgcgagagg     60

cgtttctcca tccgcggctc gcgcgctcgc tctgagcccc cgcgcccagg tggg  atg     117
                                                            Met
                                                             1

gaa gaa gcc tgt cag gtg cag aca act aag cga ggg gat cct cat gag      165
Glu Glu Ala Cys Gln Val Gln Thr Thr Lys Arg Gly Asp Pro His Glu
              5                  10                  15

tta aga aac ata ttt cta cag tat gcc agt act gag gtt gat gga gag      213
Leu Arg Asn Ile Phe Leu Gln Tyr Ala Ser Thr Glu Val Asp Gly Glu
         20                  25                  30

cgt tac atg acc cca gaa gac ttt gtt cag cgc tat ctt gga ctg tat      261
Arg Tyr Met Thr Pro Glu Asp Phe Val Gln Arg Tyr Leu Gly Leu Tyr
     35                  40                  45

aat gat cca aat agt aac cca aag atc gtg cag ctc ttg gca gga gta      309
Asn Asp Pro Asn Ser Asn Pro Lys Ile Val Gln Leu Leu Ala Gly Val
 50                  55                  60                  65

gct gat caa acc aag gat ggg ttg atc tcc tat caa gag ttt ttg gca      357
Ala Asp Gln Thr Lys Asp Gly Leu Ile Ser Tyr Gln Glu Phe Leu Ala
                 70                  75                  80

ttt gaa tct gtt tta tgt gct cca gat tcc atg ttc ata gtg gct ttc      405
Phe Glu Ser Val Leu Cys Ala Pro Asp Ser Met Phe Ile Val Ala Phe
             85                  90                  95

cag ttg ttt gac aag agt gga aat gga gag gtg aca ttt gaa aat gtc      453
Gln Leu Phe Asp Lys Ser Gly Asn Gly Glu Val Thr Phe Glu Asn Val
        100                 105                 110

aaa gaa att ttt gga cag act att att cat cat cat atc cct ttt aac      501
Lys Glu Ile Phe Gly Gln Thr Ile Ile His His His Ile Pro Phe Asn
    115                 120                 125
```

```
tgg gat tgt gaa ttt atc cga ctg cat ttt ggg cat aac cgg aag aag      549
Trp Asp Cys Glu Phe Ile Arg Leu His Phe Gly His Asn Arg Lys Lys
130                 135                 140                 145

cat ctt aac tac aca gaa ttc acg cag ttt ctc cag gag ctg caa ttg      597
His Leu Asn Tyr Thr Glu Phe Thr Gln Phe Leu Gln Glu Leu Gln Leu
                150                 155                 160

gaa cat gca aga caa gcc ttt gca ctc aaa gac aaa agc aaa agt ggc      645
Glu His Ala Arg Gln Ala Phe Ala Leu Lys Asp Lys Ser Lys Ser Gly
            165                 170                 175

atg att tct ggt ctg gat ttc agt gac atc atg gtt acc att aga tct      693
Met Ile Ser Gly Leu Asp Phe Ser Asp Ile Met Val Thr Ile Arg Ser
        180                 185                 190

cac atg ctt act cct ttt gtg gag gag aac tta gtt tca gca gct gga      741
His Met Leu Thr Pro Phe Val Glu Glu Asn Leu Val Ser Ala Ala Gly
    195                 200                 205

gga agt atc tca cac cag gtt agc ttc tcc tac ttc aat gca ttt aac      789
Gly Ser Ile Ser His Gln Val Ser Phe Ser Tyr Phe Asn Ala Phe Asn
210                 215                 220                 225

tcg tta ctg aat aac atg gag ctt gtt cgt aag ata tat agc act cta      837
Ser Leu Leu Asn Asn Met Glu Leu Val Arg Lys Ile Tyr Ser Thr Leu
                230                 235                 240

gct ggc aca agg aaa gat gtt gaa gtc aca aag gag gaa ttt gcc cag      885
Ala Gly Thr Arg Lys Asp Val Glu Val Thr Lys Glu Glu Phe Ala Gln
            245                 250                 255

agt gcc ata cgc tat gga caa gtc aca cca cta gaa att gat att cta      933
Ser Ala Ile Arg Tyr Gly Gln Val Thr Pro Leu Glu Ile Asp Ile Leu
        260                 265                 270

tat cag ctt gca gac tta tat aat gct tca ggg cgc ttg act ttg gca      981
Tyr Gln Leu Ala Asp Leu Tyr Asn Ala Ser Gly Arg Leu Thr Leu Ala
    275                 280                 285

gat att gag aga ata gcc cca ttg gct gag ggg gcc tta cct tac aac     1029
Asp Ile Glu Arg Ile Ala Pro Leu Ala Glu Gly Ala Leu Pro Tyr Asn
290                 295                 300                 305

ctg gca gaa ctt cag aga cag cag tct cct ggg tta ggc agg cct atc     1077
Leu Ala Glu Leu Gln Arg Gln Gln Ser Pro Gly Leu Gly Arg Pro Ile
                310                 315                 320

tgg ctc cag att gcc gag tct gct tac aga ttc act ctg ggc tca gtt     1125
Trp Leu Gln Ile Ala Glu Ser Ala Tyr Arg Phe Thr Leu Gly Ser Val
            325                 330                 335

gct gga gct gtg gga gcc act gca gtg tat cct ata gat ctg gtg aag     1173
Ala Gly Ala Val Gly Ala Thr Ala Val Tyr Pro Ile Asp Leu Val Lys
        340                 345                 350

acc cga atg caa aac cag cgt ggc tct ggc tct gtt gtt ggg gag cta     1221
```

```
Thr Arg Met Gln Asn Gln Arg Gly Ser Gly Ser Val Val Gly Glu Leu
    355                 360                 365

atg tac aaa aac agc ttt gac tgt ttt aag aaa gtc ttg cgt tat gag     1269
Met Tyr Lys Asn Ser Phe Asp Cys Phe Lys Lys Val Leu Arg Tyr Glu
370                 375                 380                 385

ggc ttc ttt gga ctc tac agg ggt ctg ata cca caa ctt ata ggg gtt     1317
Gly Phe Phe Gly Leu Tyr Arg Gly Leu Ile Pro Gln Leu Ile Gly Val
                390                 395                 400

gct cca gaa aag gcc att aaa ctg act gtt aat gat ttt gtt cgg gac     1365
Ala Pro Glu Lys Ala Ile Lys Leu Thr Val Asn Asp Phe Val Arg Asp
            405                 410                 415

aaa ttt acc aga aga gat ggc tct gtt cca ctt cca gca gaa gtt ctt     1413
Lys Phe Thr Arg Arg Asp Gly Ser Val Pro Leu Pro Ala Glu Val Leu
        420                 425                 430

gct gga ggc tgt gct gga ggc tct cag gtc att ttt acc aac cca ttg     1461
Ala Gly Gly Cys Ala Gly Gly Ser Gln Val Ile Phe Thr Asn Pro Leu
    435                 440                 445

gag ata gtg aag att cgt ctg caa gta gct gga gag atc acc acg gga     1509
Glu Ile Val Lys Ile Arg Leu Gln Val Ala Gly Glu Ile Thr Thr Gly
450                 455                 460                 465

ccc aga gtc agc gcc ctg aat gtg ctc cgg gac ttg gga att ttt ggt     1557
Pro Arg Val Ser Ala Leu Asn Val Leu Arg Asp Leu Gly Ile Phe Gly
                470                 475                 480

ctg tat aag ggt gcc aaa gcg tgt ttc ctc cga gac att ccc ttc tct     1605
Leu Tyr Lys Gly Ala Lys Ala Cys Phe Leu Arg Asp Ile Pro Phe Ser
            485                 490                 495

gca atc tat ttt cct gtt tat gct cat tgc aaa cta ctt ctg gct gat     1653
Ala Ile Tyr Phe Pro Val Tyr Ala His Cys Lys Leu Leu Leu Ala Asp
        500                 505                 510

gaa aat gga cac gtg gga ggt tta aat ctt ctt gca gct gga gcc atg     1701
Glu Asn Gly His Val Gly Gly Leu Asn Leu Leu Ala Ala Gly Ala Met
    515                 520                 525

gca ggt gtc cca gct gca tct ctg gtg acc cct gct gat gtc atc aag     1749
Ala Gly Val Pro Ala Ala Ser Leu Val Thr Pro Ala Asp Val Ile Lys
530                 535                 540                 545

aca aga ctg cag gtg gct gcc cgc gct ggc cag acg aca tac agt ggt     1797
Thr Arg Leu Gln Val Ala Ala Arg Ala Gly Gln Thr Thr Tyr Ser Gly
                550                 555                 560

gtc atc gac tgt ttc agg aag att ctc cgg gaa gaa ggg ccc tca gca     1845
Val Ile Asp Cys Phe Arg Lys Ile Leu Arg Glu Glu Gly Pro Ser Ala
            565                 570                 575

ttt tgg aaa ggg act gca gct cga gtg ttt cga tcc tct ccc cag ttt     1893
Phe Trp Lys Gly Thr Ala Ala Arg Val Phe Arg Ser Ser Pro Gln Phe
```

```
        580                 585                 590

ggt gtt acc ttg gtc act tat gaa ctt ctc cag cgg tgg ttt tac att     1941
Gly Val Thr Leu Val Thr Tyr Glu Leu Leu Gln Arg Trp Phe Tyr Ile
    595                 600                 605

gat ttt gga ggc ctc aaa ccc gct ggt tca gaa cca aca cct aag tca     1989
Asp Phe Gly Gly Leu Lys Pro Ala Gly Ser Glu Pro Thr Pro Lys Ser
610                 615                 620                 625

cgc att gca gac ctt cct cct gcc aac cct gat cac atc ggt gga tac     2037
Arg Ile Ala Asp Leu Pro Pro Ala Asn Pro Asp His Ile Gly Gly Tyr
                630                 635                 640

aga ctc gcc aca gcc acg ttt gca ggc atc gaa aac aaa ttt ggc ctt     2085
Arg Leu Ala Thr Ala Thr Phe Ala Gly Ile Glu Asn Lys Phe Gly Leu
            645                 650                 655

tat ctc ccg aaa ttt aag tct cct agt gtt gct gtg gtt cag cca aag     2133
Tyr Leu Pro Lys Phe Lys Ser Pro Ser Val Ala Val Val Gln Pro Lys
        660                 665                 670

gca gca gtg gca gcc act cag tga tgagacaact gttgagtgtg gcaaaatggc    2187
Ala Ala Val Ala Ala Thr Gln  *
    675                 680

gccttgaaga aagagcctag gagagcagcc ctgtaatgta tccagtcagc tgcatggtac   2247

tgactgagct gaggagtcaa actcttcttt ctgtatgaca tatacatata cttgtttata   2307

aaataatcat ttgcccaggg aaaaaaccac aacgctgttt caagctttag tcttatgtgt   2367

tgaaatgttt ttgtaagcct tggcatgaat tagtgttcta gactctgctt tgcacagctt   2427

gcacttacag tgattgtaca tattgtacat ctttgtacag agacatcttg gcacctcatc   2487

ccaacaaatc acatttgtag aaatgtaatg cggttctgag tggcttgaaa tgtacagaat   2547

gttttgaaag tgttttatta agaatcacac aaaaataaat gtattaaaat taaattcatt   2607

ctcttattgg tgacttatgg aaataaagca tcaatattgg atgtaaaaaa aaaaaa       2663
```

<210> 4
<211> 6833
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (19)..(2364)

<220>
<221> misc_feature
<222> (1)...(6833)

<223> n = a,t,c or g

<400> 4

```
cgcgaggcaa canatgac  atg ttg gcc ttc ctg tct ggg atg ccg gtg acc      51
                     Met Leu Ala Phe Leu Ser Gly Met Pro Val Thr
                       1               5                  10

aga aac acc aag tac ctc gac ctc aag aat tca caa gag atg ctc cgc      99
Arg Asn Thr Lys Tyr Leu Asp Leu Lys Asn Ser Gln Glu Met Leu Arg
             15                  20                  25

tac aaa gag gtc tgc tac tac atg ctc ttt gcc ctg gct gcc tac ggg     147
Tyr Lys Glu Val Cys Tyr Tyr Met Leu Phe Ala Leu Ala Ala Tyr Gly
         30                  35                  40

tgg ccc atg tac ctg atg cgg aag ccc gcc tgc ggc ctc tgc caa ctg     195
Trp Pro Met Tyr Leu Met Arg Lys Pro Ala Cys Gly Leu Cys Gln Leu
     45                  50                  55

gct cgg tcc tgc tcg tgt tgc ctg tgt cct gcg agg ccg cgg ttc gcc     243
Ala Arg Ser Cys Ser Cys Cys Leu Cys Pro Ala Arg Pro Arg Phe Ala
 60                  65                  70                  75

cct gga gtc acc atc gag gaa gac aac tgc tgt ggc tgt aat gcc att     291
Pro Gly Val Thr Ile Glu Glu Asp Asn Cys Cys Gly Cys Asn Ala Ile
                 80                  85                  90

gcc atc cgg cgc cac ttc ctg gac gag aac atg act gcg gtg gac atc     339
Ala Ile Arg Arg His Phe Leu Asp Glu Asn Met Thr Ala Val Asp Ile
             95                  100                 105

gtc tat acc tcc tgc cat gat gcg gtc tat gaa acg ccc ttc tac gtg     387
Val Tyr Thr Ser Cys His Asp Ala Val Tyr Glu Thr Pro Phe Tyr Val
         110                 115                 120

gcg gtg gac cat gac aag aag aaa gtg gtg atc agt atc cgg ggg acc     435
Ala Val Asp His Asp Lys Lys Lys Val Val Ile Ser Ile Arg Gly Thr
     125                 130                 135

ctg tcc ccc aag gat gcc ctg act gac ctg acg ggt gat gct gag cgc     483
Leu Ser Pro Lys Asp Ala Leu Thr Asp Leu Thr Gly Asp Ala Glu Arg
140                 145                 150                 155

ctc ccc gtg gag ggg cac cac ggc acc tgg ctg ggc cac aag ggt atg     531
Leu Pro Val Glu Gly His His Gly Thr Trp Leu Gly His Lys Gly Met
                160                 165                 170

gtc ctc tca gct gag tac atc aag aag aaa ctg gag cag gag atg gtc     579
Val Leu Ser Ala Glu Tyr Ile Lys Lys Lys Leu Glu Gln Glu Met Val
            175                 180                 185

ctg tcc cag gcc ttt ggg cga gac ctg ggc cgc gga acc aaa cac tac     627
Leu Ser Gln Ala Phe Gly Arg Asp Leu Gly Arg Gly Thr Lys His Tyr
        190                 195                 200

ggc ctg att gtg gtg ggc cac tcc ctg ggc gcg ggc act gct gcc atc     675
Gly Leu Ile Val Val Gly His Ser Leu Gly Ala Gly Thr Ala Ala Ile
```

```
    205                 210                 215

ctc tcc ttc ctt ctg cgc cca cag tat ccg acc ctc aag tgc ttt gcc      723
Leu Ser Phe Leu Leu Arg Pro Gln Tyr Pro Thr Leu Lys Cys Phe Ala
220                 225                 230                 235

tac tcc ccg cca ggg ggc ctg ctg agt gag gat gcg atg gag tat tcc      771
Tyr Ser Pro Pro Gly Gly Leu Leu Ser Glu Asp Ala Met Glu Tyr Ser
                240                 245                 250

aag gag ttc gtg act gct gtg gtt ctg ggc aaa gac ctc gtc ccc agg      819
Lys Glu Phe Val Thr Ala Val Val Leu Gly Lys Asp Leu Val Pro Arg
            255                 260                 265

att ggc ctc tct cag ctg gaa ggc ttc cgc aga cag ctc ctg gat gtc      867
Ile Gly Leu Ser Gln Leu Glu Gly Phe Arg Arg Gln Leu Leu Asp Val
        270                 275                 280

ctg cag cga agc acc aag ccc aaa tgg cgg atc atc gtg ggg gcc acc      915
Leu Gln Arg Ser Thr Lys Pro Lys Trp Arg Ile Ile Val Gly Ala Thr
    285                 290                 295

aaa tgc atc ccc aag tcg gag ctg cct gag gag gta gag gtg acc acc      963
Lys Cys Ile Pro Lys Ser Glu Leu Pro Glu Glu Val Glu Val Thr Thr
300                 305                 310                 315

ctg gcc agc acg cgg ctc tgg acc cac ccc agc gac cta act ata gcc     1011
Leu Ala Ser Thr Arg Leu Trp Thr His Pro Ser Asp Leu Thr Ile Ala
                320                 325                 330

ctc tca gcc agc act cca ctc tac ccg ccc ggc cgc atc atc cac gtg     1059
Leu Ser Ala Ser Thr Pro Leu Tyr Pro Pro Gly Arg Ile Ile His Val
            335                 340                 345

gtc cac aac cac cct gca gag cag tgc tgc tgc tgt gag cag gag gag     1107
Val His Asn His Pro Ala Glu Gln Cys Cys Cys Cys Glu Gln Glu Glu
        350                 355                 360

ccc aca tac ttt gcc atc tgg ggc gac aac aag gcc ttc aat gag gtg     1155
Pro Thr Tyr Phe Ala Ile Trp Gly Asp Asn Lys Ala Phe Asn Glu Val
    365                 370                 375

atc atc tcg cca gcc atg ctg cat gag cac ctg ccc tat gtg gtc atg     1203
Ile Ile Ser Pro Ala Met Leu His Glu His Leu Pro Tyr Val Val Met
380                 385                 390                 395

gag ggg ctc aac aag gtg ctg gag aac tac aac aag ggg aag acc gct     1251
Glu Gly Leu Asn Lys Val Leu Glu Asn Tyr Asn Lys Gly Lys Thr Ala
                400                 405                 410

ctg ctc tct gca gcc aag gtc atg gtg agc cct acc gag gtg gac ctg     1299
Leu Leu Ser Ala Ala Lys Val Met Val Ser Pro Thr Glu Val Asp Leu
            415                 420                 425

act cct gag ctc atc ttc cag cag cag cca ctc ccc acg ggg ccg ccc     1347
Thr Pro Glu Leu Ile Phe Gln Gln Gln Pro Leu Pro Thr Gly Pro Pro
        430                 435                 440
```

```
atg ccc act ggc ctt gcc ctg gag ctg ccg act gca gac cac cgc aac     1395
Met Pro Thr Gly Leu Ala Leu Glu Leu Pro Thr Ala Asp His Arg Asn
    445                 450                 455

agc agc gtc agg agc aag tcc cag tct gag atg agc ctg gag ggc ttc     1443
Ser Ser Val Arg Ser Lys Ser Gln Ser Glu Met Ser Leu Glu Gly Phe
460                 465                 470                 475

tcg gag ggg cgg ctg ctg tcg cca gtg gtt gcg gcg gcg gcc cgc cag     1491
Ser Glu Gly Arg Leu Leu Ser Pro Val Val Ala Ala Ala Arg Gln
                480                 485                 490

gac ccg gtg gag ctg ctg ctg ctg tct acc cag gag cgg ctg gca gcg     1539
Asp Pro Val Glu Leu Leu Leu Leu Ser Thr Gln Glu Arg Leu Ala Ala
            495                 500                 505

gag ctg cag gcc cgg cgg gca cca ctg gcc acc atg gag agc ctc tcg     1587
Glu Leu Gln Ala Arg Arg Ala Pro Leu Ala Thr Met Glu Ser Leu Ser
        510                 515                 520

gac act gag tcc ctg tac agc ttc gac tcg cgc cgc tcc tca ggc ttc     1635
Asp Thr Glu Ser Leu Tyr Ser Phe Asp Ser Arg Arg Ser Ser Gly Phe
    525                 530                 535

cgc agc atc cgg ggc tcc ccc agc ctc cac gct gtg ctg gag cgt gat     1683
Arg Ser Ile Arg Gly Ser Pro Ser Leu His Ala Val Leu Glu Arg Asp
540                 545                 550                 555

gaa ggc cac ctc ttc tac att gac cct gcc atc ccc gag gaa aac cca     1731
Glu Gly His Leu Phe Tyr Ile Asp Pro Ala Ile Pro Glu Glu Asn Pro
                560                 565                 570

tcc ctg agc tcg cgc act gag ctg ctg gcg gcc gac agc ctg tcc aag     1779
Ser Leu Ser Ser Arg Thr Glu Leu Leu Ala Ala Asp Ser Leu Ser Lys
            575                 580                 585

cac tca cag gac acg cag ccc ctg gag gcg gcc ctg ggc agt ggc ggc     1827
His Ser Gln Asp Thr Gln Pro Leu Glu Ala Ala Leu Gly Ser Gly Gly
        590                 595                 600

gtc act cct gag cgg ccc ccc agt gct gcg gcc aat gac gag gag gaa     1875
Val Thr Pro Glu Arg Pro Pro Ser Ala Ala Ala Asn Asp Glu Glu Glu
    605                 610                 615

gag gtt ggc ggt ggg ggt ggc ggg ccg gcc tcc cgc ggg gag ctg gcg     1923
Glu Val Gly Gly Gly Gly Gly Gly Pro Ala Ser Arg Gly Glu Leu Ala
620                 625                 630                 635

ctg cac aat ggg cgc ctg ggg gac tcg ccc agt cct cag gtg ctg gaa     1971
Leu His Asn Gly Arg Leu Gly Asp Ser Pro Ser Pro Gln Val Leu Glu
                640                 645                 650

ttc gcc gag ttc atc gac agc ctc ttc aac ctg gac agc aag agc agc     2019
Phe Ala Glu Phe Ile Asp Ser Leu Phe Asn Leu Asp Ser Lys Ser Ser
            655                 660                 665
```

```
tcc ttc caa gac ctc tac tgc atg gtg gtg ccc gag agc ccc acc agt     2067
Ser Phe Gln Asp Leu Tyr Cys Met Val Val Pro Glu Ser Pro Thr Ser
        670                 675                 680

gac tac gct gag ggc ccc aag tcc ccc agc cag caa gag atc ctg ctc     2115
Asp Tyr Ala Glu Gly Pro Lys Ser Pro Ser Gln Gln Glu Ile Leu Leu
    685                 690                 695

cgt gcc cag ttc gag ccc aac ctg gtg ccc aag ccc cca cgg ctc ttt     2163
Arg Ala Gln Phe Glu Pro Asn Leu Val Pro Lys Pro Pro Arg Leu Phe
700                 705                 710                 715

gcc ggc tca gcc gac ccc tcc tcg ggc atc tca ctc tcg ccc tcc ttc     2211
Ala Gly Ser Ala Asp Pro Ser Ser Gly Ile Ser Leu Ser Pro Ser Phe
                720                 725                 730

ccg ctc agc tcc tcg ggt gag ctc atg gac ctg acg ccc acg ggc ctc     2259
Pro Leu Ser Ser Ser Gly Glu Leu Met Asp Leu Thr Pro Thr Gly Leu
            735                 740                 745

agt agc cag gaa tgc ctg gcg gct gac aag atc cgg act tct acc ccc     2307
Ser Ser Gln Glu Cys Leu Ala Ala Asp Lys Ile Arg Thr Ser Thr Pro
        750                 755                 760

act ggc cac gga gcc agc ccc gcc aag caa gat gag ctg gtc atc tca     2355
Thr Gly His Gly Ala Ser Pro Ala Lys Gln Asp Glu Leu Val Ile Ser
    765                 770                 775

gca cgc tag cacccca gttgcgtggc cagccgggcc caggcaggag caggtggccc     2411
Ala Arg  *
780

tgtgggcacc tggtgcctgc cccctgccgg gcagctttaa ggacagaccc ccaggggcag   2471

tttagcctca ggcacaggca tcgctgctga gctgggggtc cgcatcccta cctcagctta   2531

ggacccccag agccaaggtg gctgggatct ggccccacag atggggaaag atggggaagg   2591

gtgtggagtg gggaggagcc tgggcagcct gctgggtggg ccacactcag cctgactgcc   2651

ctccatgggg gcattctggc acccccctgct ccaggacagg ccatgggcaa gctgcctccc   2711

atcactgcct gctggctgct ctcccagggg ccaggtggag agcagtgccc cccgacacat   2771

gtattctcat ctgtggtcca ggccggcatc gtcctggcca ccccccagat ctggtgcctg   2831

ctggccggcc ccctggggtg cccctgccga ggtggcctgc agtgctgtac atgtttacag   2891

aagctgctgg gcttggctca ggatgtgttc tgggcttgca agcccccccgc ccaatcatgt   2951

gttcagtagc catcctctga gcagggccca aggcagccag gggcctggag gggccagagg   3011

agggtggggt cagggccgcc ccttctctgc cttgtgcctc tcatgctgcc tcctctgccc   3071

atgggtcctg ggcacccagg cctgccctgc ctgctggcta cttcctggct taccttctac   3131

ccccaaggat cctcaccacc caaagggtgg tgggcactgc tgtgaccacc ccagctgcag   3191
```

agtcagtgcc ctgggtggaa ggaaggcact gagagccccc ttcctctgag ggccccacct 3251

caccccttgg tgtcaccccc accacgccta ggcagctctg ggccctggga tctggaacca 3311

acacacccct gttcccctca gctttccctc ctcgctggcc tgggcaccct cctgggagca 3371

ggccttcctc cctcccaccc ccaatgtcct gttggtagga ggtggggcca agagtggggt 3431

atggtgggcc ttggctggag acctctgtcc actgcccagg gaggggcctg gggctgggag 3491

cagtcccggt ttagcctgag gtccccatag ggcttcctcc cctgctgggt ttgggaagca 3551

gttagggaga tagcgacccg gagtttcccc agaagcgggg tgggagggtg tgcatgctag 3611

tgttggcgcg tatgcatgtg catgagtgtg caccgttcct aaggaagggg cctctggggc 3671

tgcccaccct acctgccctg cctgcctgct gcccctccca gcctgccaag aaaacggtag 3731

gggagcatga tggggccttt gaggcagggt cgcagggaca agctcagctt taggcaccat 3791

ctgttcccat cgcgcctgct gctgtgaccc gttttggaaa actggtgtgt accgaggcgc 3851

tgactgcacg gctgaccgcc tgctcgtgcc ttcattctgc agcggcatgg tccctcccat 3911

tctggctcca cctgcagcct ccctgggtgg cctaggctcc cccgaccaag agacctccct 3971

ctcatgatca ctggtacctg ggggcctgaa ttctggcccc cggctcccca cacagctggg 4031

actggcctgg atggctgtcc tgggagcccc tgcccaccct gacagaggga gctgggcctc 4091

ccctcatcct ctgtaactcc cgccttcacc agactcaagg acaccctggc cctgctgagg 4151

catacagagc ttcagcccag cacagaagca agacaaaatc agtggctctt agagtttaga 4211

aaacaagaca gactctcaga tgaaagatct gacaagcacc gtggccagtc acagggagag 4271

acttgatgtc tggcctttta attcctcctc tgccagggtg ggtcctggga cctctaatgt 4331

gggcatgtcg tccaccccag gacaagccat cagggacaga ccccccaccc ccaaggctgc 4391

agccacacca tgtttcaggc ttggggctgg ggcaggcttg ggctcaatcc tgggcaccca 4451

ggggcagccc acccctaacc tggctcctac ccaccttgcc cttgaaggat gggcctgctg 4511

cacgtctccc tcctccaccc cataccacac tgggggggtct gagccacccc cctcagcccc 4571

gttcggctca gaccgacccc cactccatcc ccagacctgc agcacaagtg cgcgggcctg 4631

tcctcccagg ggcctgggcg actccatatg caatcagtag cgagcagccg ggccccacag 4691

accctcatgc actctcttac gtgccattct ccccagactt tttttgtact taatgtatga 4751

aagatccaaa ctaatattgc tgtaaaaagg agagacaaat taatatagct tattctataa 4811

atatatctgt atataaaggt ttctgtatat tgtatagagc tgtgtataaa ctggatgtag 4871

aagcacgctg gctgcctcga atgtctttgc atcaggtggg gactgggtag aaatttgatg 4931

tcgaggttgc agcagagcag ggggttggca tggggccggg ggccgggggg gcctctgcca 4991

ctgttccctg gatggaacag aaagcctgct cctgctcggc tagtgccctg gcccggggcc 5051

acactttgag ttgcagggag ggagtagaag acccttgaag ccccctgtga atgagggcag 5111

ccacttgcgg agtcctgccc cactttgagc cctcctcttc ctctgcgaaa tggcctgatg 5171

ccagtgctgt gtgggtccac aggaggccag gagagtccca cggtggggag gacagggctg 5231

taccttctcc tgggctggcc tctacgccca ttacccatta accctcaggt gccagcatcc 5291

ctctccccag tgctgccttc ggtctaccac ctcctcctgg ccctgctccc actcaaggga 5351

cagtgatggg tgctgagagc tgattggaac tggagagggc accatttact gatcactgac 5411

ctggcacttt acctccactg taaggcaggg atactgagtg tgctttatag atgaggcctc 5471

tgaggacaga gggcaggcct tggggactag gtggagctgg ctacagggga cagccatgtc 5531

tgctgggcta gggctgaaat ctcagcccct cactcactgt ggctcggtag agaagccagg 5591

ggcacagatg aggactcatc tccattgatg ggcccccta ggtccttgta tgcaagtccc 5651

ctgggctact ttcaccggcc cagccacctt cctgccccag cctctgcccc agcgctgctt 5711

gggacccaac ttcattatgg agttggcaga tggcagcctc aactcttggc tgagcccttg 5771

agtctgggac atttcagcca cctctttcct ccagtccaga gatgaaaatc cctggggaca 5831

gttgctcctt tgctcagtga cctagtgtaa caggggagat ggcagggcct gagctccctt 5891

agccaggtca gctgctacag ggttaacagg aggctccatt ccacccccttc caacttcaag 5951

gctaccctag agattgaata atctatactc ttaattgatt ataatgcaat ggagttgggg 6011

cgttagggac aaagtacgag tcttcccttc tacctccagc ccttgctgac caggacaggg 6071

acaatgtgta gctcaacgga tggtggtgag aacttagatg atggtcagtt atgcagtatg 6131

tgggatacgg aggaaagatc cgtgggtatg tggaggcttg tagagaagct ggttctgtgg 6191

ctggtcccag gcgactcgta atgtaaatcc gtttctcaga atcgcgtggt gtaggcgggt 6251

gtctactttg tccgcaggca ggcctgaccc cgggtggagg aggggcaggg tggagaataa 6311

caattgtctt aagggagtct gcaagacagg agggggtggc agagaagaac ctcagctctg 6371

aagaagctca ctgcccagcc cttcccacct tcctcttcac ggacctagca ccttcctggg 6431

cctcagtttc tctcattgcc ttgggtgctg gagtatgtgg ggcctcctct cctatctcca 6491

ggccttcagc cccggctgcc acggggtgtg ggtacctctt ggttgggtct cggggtagga 6551

tgatgtaatg gttctgtgca ttcgccagcg agggcagctg gggtctgttc ctagctctcc 6611

```
tgcttaccca cagtgcttct cttggccgta tcagggcacc gctgtgcctc cgctttctca   6671

tcttgaagat tactggtccc cagggtaggt cagtgcccct aagcttaggg ggcttgttga   6731

gcatgttctg tggttctgtg tgcaaggcct gaaccatgac agctctggcc cagcgtggcc   6791

tggtcctggt ccctggcaca tccagtgggg cccgccccac ct                      6833
```

<210> 5
<211> 1720
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (197)..(619)

<400> 5

```
atgatcgcct tcggaaccgg cccggaattc ccgggtcgac ccacgcgtcc gctctttggc     60

actagttcag aatggtgatg tgtcggcccc ctctgccata ctcagaacac cagaaagcac    120

aaaaccgggt cctgtttgtc agccaccagt gagtcagagc cgctccctgt tttcttctgt    180

cccgtccaag ccacca    atg tct ctg gag cct caa aat ggg acg tat gca      229
                     Met Ser Leu Glu Pro Gln Asn Gly Thr Tyr Ala
                       1               5                  10

gga cca gcg cca gca ttc cag cca ttt ttc ttc act gga gca ttt cca      277
Gly Pro Ala Pro Ala Phe Gln Pro Phe Phe Phe Thr Gly Ala Phe Pro
             15                  20                  25

ttt aat atg caa gag ctg gta ctc aag gtg aga att cag aac cca tct      325
Phe Asn Met Gln Glu Leu Val Leu Lys Val Arg Ile Gln Asn Pro Ser
         30                  35                  40

ctt cga gaa aat gat ttc att gaa att gaa ctg gac cga cag gag ctc      373
Leu Arg Glu Asn Asp Phe Ile Glu Ile Glu Leu Asp Arg Gln Glu Leu
     45                  50                  55

acc tac caa gag ttg ctc aaa gtg tgt tgc tgt gag ctg ggt gtt aat      421
Thr Tyr Gln Glu Leu Leu Lys Val Cys Cys Cys Glu Leu Gly Val Asn
 60                  65                  70                  75

cca gat caa gtg gag aag atc aga aag tta ccc aat act ctg tta agg      469
Pro Asp Gln Val Glu Lys Ile Arg Lys Leu Pro Asn Thr Leu Leu Arg
                 80                  85                  90

aag gac aag gat gtt gct cga ctc caa gat ttc cag gag ctg gaa ctg      517
Lys Asp Lys Asp Val Ala Arg Leu Gln Asp Phe Gln Glu Leu Glu Leu
             95                  100                 105

gtt ctg atg ata agt gaa aat aat ttt ctg ttc aga aat gct gca tcc      565
```

```
Val Leu Met Ile Ser Glu Asn Asn Phe Leu Phe Arg Asn Ala Ala Ser
        110                 115                 120

aca ctg act gaa agg cct tgc tat aac agg aga gct tca aaa ctg act      613
Thr Leu Thr Glu Arg Pro Cys Tyr Asn Arg Arg Ala Ser Lys Leu Thr
    125                 130                 135

tac taa tgcagcaggg acttttatca ctgagtatta tgacagtgtg catcacctct      669
Tyr  *
140

gggccaagga caagccattg atctaaatgc ctcagatgcc cgggagggcc tctggtgcca    729

ctgcatagta tatactaaca tcattctgcc aaggtaggaa gcccctgacc cccaagcagt    789

ggtgccactc ttccaagcct cttggtgcac agtaaaccta ttgcttgaag ctttgaacag    849

ctgagaagtg gtctggagag gcagaagctg aaggttctat atccagtgtg ttttatgtcc    909

agaatgtaag agagttgtct aagcagcagc tgagagagag cggagcctat ttctagccac    969

tcctgttgac agtgcacctg aagggctggg atgcgttttt cttggtgttg catgctcaca   1029

actctgctga cattgggaac ttatgagaga ggaagactcg ggaaagcaca gatactggac   1089

agatggattc tggtgtgggg aaagcacaga tactggacag atggattcta gtgtgacttg   1149

tgactgtgag gtttcctata acatatttat aaatgttact caggttaaaa gtatttaaga   1209

atacagttaa ctaattgtaa atatgctgtt aaccaaaaga gctttccctc cctcactttt   1269

tcctttgtaa acactcatga ctgcttctct gtctcgagtc atctctgcat taactcccct   1329

tcgtggtcac tagagggctc tctgatgcct tctaaaagaa caactgcttt tttacaatgc   1389

cccccacccc caccccgccc atagagacag ggtctcacta tgtggcccag gctggtctca   1449

aacttctggc cttaagtgat gctcctgtcc ttggcctccc aaagtgctgg gattacaggt   1509

gtgagccact gcacccagcc cactttttt ttactctgaa gtgattccag tcatatgtgt   1569

gtgtaactgc atcctatgag aagagcacaa atattgctgt tccatgttct ccactttcat   1629

tttccactac aaatgaaaag caatttttga gactgaatct gttgctattt taaaggttat   1689

tgtgggaaac tgagctaaag gagttagcat c                                  1720
```

<210> 6
<211> 1658
<212> DNA
<213> Homo sapiens

<220>
<221> CDS

<222> (523)..(1479)

<400> 6

```
ctagtacgcc tgtggaacgc ctgcaggtac cggtccggaa ttcccgggtc gacccacgcg      60

tccgcccacg cgtccgctct gtctcaaata ataataataa taataataat aataataata     120

ataataataa tgtaggggac ttgatgaagg gaaaggatca gatagattct gaaaagaagt     180

acagaacctg ctaaggccat catacctatt gatcggaagt cagtccatca gatttgctct     240

ggaccggtgg tactgagtct aatcactgcg tgtaagaaga tagtaggaaa cattctggat     300

gctaggtgcc agtaatattg atctaaagct taaggactat ggaatggatc tcattgaagt     360

ttcaggcaat ggatgtgggg tagaagaaga aaacttcgaa ggcttaatct ctttcagctc     420

tgaaacatca cacatctaag attcaagagt ttgccgacct aactcgggtt gaaacttttg     480

gctttcgggg gaaagctctg agctcacttt gtgcactgag tg   atg tca cca ttt      534
                                                     Met Ser Pro Phe
                                                       1

cta cct gcc acg tat cgg cga agg ttg gga ctc gac tgg tgt ttg atc      582
Leu Pro Ala Thr Tyr Arg Arg Arg Leu Gly Leu Asp Trp Cys Leu Ile
  5                  10                  15                  20

acg atg gga aaa tca tcc aga aaa ccc cct acc ccc acc cca gag gga      630
Thr Met Gly Lys Ser Ser Arg Lys Pro Pro Thr Pro Thr Pro Glu Gly
                 25                  30                  35

ccc aca gtc agc gtg aag cag tta ttt tct acg cta cct gtg cgc cat      678
Pro Thr Val Ser Val Lys Gln Leu Phe Ser Thr Leu Pro Val Arg His
             40                  45                  50

aag gaa ttt caa agg aat att aag aag aaa cgt gcc tgc ttc ccc ttc      726
Lys Glu Phe Gln Arg Asn Ile Lys Lys Lys Arg Ala Cys Phe Pro Phe
         55                  60                  65

gcc ttc tgc cgt gat tgt cag ttt ctt gag ggc tcc cca gcc atg ctt      774
Ala Phe Cys Arg Asp Cys Gln Phe Leu Glu Gly Ser Pro Ala Met Leu
     70                  75                  80

cct gta cag cct gca aaa ctt aca gaa cct gct aag gcc atc aaa cct      822
Pro Val Gln Pro Ala Lys Leu Thr Glu Pro Ala Lys Ala Ile Lys Pro
 85                  90                  95                  100

att gat cgg aag tca gtc cat cag att tgc tct ggg ccg gtg gta ctg      870
Ile Asp Arg Lys Ser Val His Gln Ile Cys Ser Gly Pro Val Val Leu
                105                 110                 115

agt cta agc act gcg gtg aag aag ata gta gga aac agt ctg gat gct      918
Ser Leu Ser Thr Ala Val Lys Lys Ile Val Gly Asn Ser Leu Asp Ala
            120                 125                 130

ggt gcc act aat att gat cta aag ctt aag gac tat gga atg gat ctc      966
Gly Ala Thr Asn Ile Asp Leu Lys Leu Lys Asp Tyr Gly Met Asp Leu
```

```
        135                 140                 145

att gaa gtt tca ggc aat gga tgt ggg gta gaa gaa gaa aac ttc gaa     1014
Ile Glu Val Ser Gly Asn Gly Cys Gly Val Glu Glu Glu Asn Phe Glu
    150                 155                 160

ggc tta tct ctt tca gct ctg aaa cat cac aca tct aag att cga gag     1062
Gly Leu Ser Leu Ser Ala Leu Lys His His Thr Ser Lys Ile Arg Glu
165                 170                 175                 180

ttt gcc gac cta act cgg gtt gaa act ttt ggc ttt cag ggg aaa gct     1110
Phe Ala Asp Leu Thr Arg Val Glu Thr Phe Gly Phe Gln Gly Lys Ala
                185                 190                 195

ctg agc tca ctt tgt gca ctg agt gat gtc acc att tct acc tgc cac     1158
Leu Ser Ser Leu Cys Ala Leu Ser Asp Val Thr Ile Ser Thr Cys His
            200                 205                 210

gta tcg gcg aag gtt ggg act cga ctg gtg ttt gat cac gat ggg aaa     1206
Val Ser Ala Lys Val Gly Thr Arg Leu Val Phe Asp His Asp Gly Lys
        215                 220                 225

atc atc aag aaa acc ccc tac ccc cac ccc aga ggg acc aca gtc agc     1254
Ile Ile Lys Lys Thr Pro Tyr Pro His Pro Arg Gly Thr Thr Val Ser
    230                 235                 240

gtg aag cag tta ttt tct acg cta cct gtg cgc cat aag gaa ttt caa     1302
Val Lys Gln Leu Phe Ser Thr Leu Pro Val Arg His Lys Glu Phe Gln
245                 250                 255                 260

agg aat att aag aag aaa cgt gcc tgc ttc ccc ttc gcc ttc tgc cgt     1350
Arg Asn Ile Lys Lys Lys Arg Ala Cys Phe Pro Phe Ala Phe Cys Arg
                265                 270                 275

gat tgt cag ttt ctt gag ggc tcc cca gcc atg ctt cct gta cag cct     1398
Asp Cys Gln Phe Leu Glu Gly Ser Pro Ala Met Leu Pro Val Gln Pro
            280                 285                 290

gca aaa ctg act gta act gga gag cta cgg gca tgc aga agt tgg aag     1446
Ala Lys Leu Thr Val Thr Gly Glu Leu Arg Ala Cys Arg Ser Trp Lys
        295                 300                 305

acg agg gaa ggc atc aca gag gct gtg ggg tga accgactt caaggaatgg     1497
Thr Arg Glu Gly Ile Thr Glu Ala Val Gly  *
    310                 315

gtccttccct tcagaaccac atgtgtgcgg gacacccaga cagaaaacac aaatgcaaag   1557

tcaagtggag ggcatttgga aggagcagtg aagccaagcc aggaaacacc aagatggcga   1617

gccagtgtgg ttgtagagat tgcagagagg gacaagataa g                       1658
```

<210> 7
<211> 3739

<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (460)..(3621)

<400> 7

```
gtctgctgta ataccctcta ctatagggac cactttgtac aagaaagctg ggtacgcgta     60

agcttgggcc cctcgaggga ttctctagag cgtccgcggg ggctgccagg gtatttcggg    120

aaggggcgt gaggaggcgg cggcggcagc ggcgggtagg gcaggcagca gagggaagga    180

gaaagaaagg aaggaagagg gcggggagtc ctcagaggag gaggcgggac cggccgggca    240

cctcacctct cggcggcggc ggcggcggcg gcggcggcgg cgggcggccg gggaggcggc    300

gccgcccccc gcccgcggtc ttctcagcgc agtcgggcgc ggacccgctg gtcccgggca    360

gcggccaagg ctactggggc gggagcagtg ggccggtcgg cggcggcagc ggcagcggcg    420

gaggaggagg aggctggagt gggcgcggag gcgaccgcc   atg gcg ttc ctc aaa     474
                                              Met Ala Phe Leu Lys
                                                1               5

ctc cgt gac cag cca tca ctg gtg caa gct ata ttt aac gga gat cct      522
Leu Arg Asp Gln Pro Ser Leu Val Gln Ala Ile Phe Asn Gly Asp Pro
                10                  15                  20

gat gaa gtt cga gca cta ata ttt aag aaa gaa gat gtt aac ttt cag      570
Asp Glu Val Arg Ala Leu Ile Phe Lys Lys Glu Asp Val Asn Phe Gln
            25                  30                  35

gac aat gaa aag cga acc cca ttg cac gcc gca gct tac ctt gga gat      618
Asp Asn Glu Lys Arg Thr Pro Leu His Ala Ala Ala Tyr Leu Gly Asp
        40                  45                  50

gca gaa atc att gaa ctt ctt att tta tct gga gct aga gtt aat gcc      666
Ala Glu Ile Ile Glu Leu Leu Ile Leu Ser Gly Ala Arg Val Asn Ala
    55                  60                  65

aaa gac agc aaa tgg ttg aca cct tta cac aga gca gtt gca tct tgt      714
Lys Asp Ser Lys Trp Leu Thr Pro Leu His Arg Ala Val Ala Ser Cys
70                  75                  80                  85

agt gag gaa gca gtt cag gta ctt ttg aag cat tct gca gat gtt aat      762
Ser Glu Glu Ala Val Gln Val Leu Leu Lys His Ser Ala Asp Val Asn
                90                  95                  100

gct cga gac aaa aat tgg caa acc cct tta cat ata gct gct gct aat      810
Ala Arg Asp Lys Asn Trp Gln Thr Pro Leu His Ile Ala Ala Ala Asn
            105                 110                 115

aaa gct gta aag tgt gct gaa gct ttg gta cct ctt ctg agt aat gta      858
Lys Ala Val Lys Cys Ala Glu Ala Leu Val Pro Leu Leu Ser Asn Val
        120                 125                 130
```

```
aac gta tct gat cga gca ggg agg act gca tta cat cat gca gct ttc      906
Asn Val Ser Asp Arg Ala Gly Arg Thr Ala Leu His His Ala Ala Phe
    135                 140                 145

agt gga cat ggt gag atg gtc aaa cta ctc ttg tct aga ggt gcc aat      954
Ser Gly His Gly Glu Met Val Lys Leu Leu Leu Ser Arg Gly Ala Asn
150                 155                 160                 165

att aat gct ttt gac aag aaa gat agg cgt gct atc cat tgg gca gca     1002
Ile Asn Ala Phe Asp Lys Lys Asp Arg Arg Ala Ile His Trp Ala Ala
                170                 175                 180

tat atg ggt cac att gaa gta gtg aaa ttg ctt gtg tcg cat gga gct     1050
Tyr Met Gly His Ile Glu Val Val Lys Leu Leu Val Ser His Gly Ala
            185                 190                 195

gaa gtg aca tgc aag gat aaa aag tct tat aca cct ctt cat gca gca     1098
Glu Val Thr Cys Lys Asp Lys Lys Ser Tyr Thr Pro Leu His Ala Ala
        200                 205                 210

gcc tct agt gga atg atc agc gta gtc aag tac ctt cta gat ctt gga     1146
Ala Ser Ser Gly Met Ile Ser Val Val Lys Tyr Leu Leu Asp Leu Gly
    215                 220                 225

gtt gat atg aat gaa cca aat gcc tat gga aat aca cct ctt cat gta     1194
Val Asp Met Asn Glu Pro Asn Ala Tyr Gly Asn Thr Pro Leu His Val
230                 235                 240                 245

gcc tgc tat aat gga caa gat gtt gta gtg aat gaa ctt ata gac tgt     1242
Ala Cys Tyr Asn Gly Gln Asp Val Val Val Asn Glu Leu Ile Asp Cys
                250                 255                 260

ggt gct att gtg aat caa aag aat gaa aaa gga ttt act cct ttg cac     1290
Gly Ala Ile Val Asn Gln Lys Asn Glu Lys Gly Phe Thr Pro Leu His
            265                 270                 275

ttt gct gct gca tca aca cat gga gca ttg tgt tta gag ctt cta gtt     1338
Phe Ala Ala Ala Ser Thr His Gly Ala Leu Cys Leu Glu Leu Leu Val
        280                 285                 290

ggc aat ggg gcc gat gtc aat atg aag agt aaa gat ggg aaa acc cca     1386
Gly Asn Gly Ala Asp Val Asn Met Lys Ser Lys Asp Gly Lys Thr Pro
    295                 300                 305

cta cac atg act gct ctc cac ggt aga ttc tcc cga tca caa acc att     1434
Leu His Met Thr Ala Leu His Gly Arg Phe Ser Arg Ser Gln Thr Ile
310                 315                 320                 325

atc cag agt gga gct gta atc gac tgt gag gat aag aat gga aat acc     1482
Ile Gln Ser Gly Ala Val Ile Asp Cys Glu Asp Lys Asn Gly Asn Thr
                330                 335                 340

cct ttg cac ata gca gca cgg tat ggc cat gag ctg ctg atc aac act     1530
Pro Leu His Ile Ala Ala Arg Tyr Gly His Glu Leu Leu Ile Asn Thr
            345                 350                 355
```

```
ctt att aca agt ggt gct gac act gca aag cgt ggc ata cat gga atg      1578
Leu Ile Thr Ser Gly Ala Asp Thr Ala Lys Arg Gly Ile His Gly Met
        360                 365                 370

ttc ccc ctc cat ttg gca gcc tta agc ggc ttt tca gat tgc tgc aga      1626
Phe Pro Leu His Leu Ala Ala Leu Ser Gly Phe Ser Asp Cys Cys Arg
    375                 380                 385

aaa ctt ctt tct tca gga ttt gat ata gat acc cca gat gat ttt ggc      1674
Lys Leu Leu Ser Ser Gly Phe Asp Ile Asp Thr Pro Asp Asp Phe Gly
390                 395                 400                 405

agg act tgt cta cat gca gct gca gct gga ggg aat ttg gag tgc cta      1722
Arg Thr Cys Leu His Ala Ala Ala Ala Gly Gly Asn Leu Glu Cys Leu
                410                 415                 420

aac ctt ctg ctg aat act ggt gca gac ttt aat aaa aag gac aaa ttt      1770
Asn Leu Leu Leu Asn Thr Gly Ala Asp Phe Asn Lys Lys Asp Lys Phe
            425                 430                 435

ggg aga tct cca ctg cac tac gct gct gcc aac tgc aat tac cag tgc      1818
Gly Arg Ser Pro Leu His Tyr Ala Ala Ala Asn Cys Asn Tyr Gln Cys
        440                 445                 450

ctg ttt gct ctt gtg gga tca gga gca agt gtg aat gac ctt gat gaa      1866
Leu Phe Ala Leu Val Gly Ser Gly Ala Ser Val Asn Asp Leu Asp Glu
    455                 460                 465

aga ggc tgc aca ccc ctg cac tat gca gct aca tca gac aca gat ggc      1914
Arg Gly Cys Thr Pro Leu His Tyr Ala Ala Thr Ser Asp Thr Asp Gly
470                 475                 480                 485

aag tgc ctg gaa tac tta tta aga aac gat gca aat cca ggg atc cgt      1962
Lys Cys Leu Glu Tyr Leu Leu Arg Asn Asp Ala Asn Pro Gly Ile Arg
                490                 495                 500

gat aag caa gga tac aac gca gtt cat tat tca gct gct tat ggt cac      2010
Asp Lys Gln Gly Tyr Asn Ala Val His Tyr Ser Ala Ala Tyr Gly His
            505                 510                 515

cgt cta tgt ctt cag ctg att gca agt gaa act cct cta gat gtt tta      2058
Arg Leu Cys Leu Gln Leu Ile Ala Ser Glu Thr Pro Leu Asp Val Leu
        520                 525                 530

atg gaa acc tca gga aca gac atg ctg agt gat tca gat aat aga gca      2106
Met Glu Thr Ser Gly Thr Asp Met Leu Ser Asp Ser Asp Asn Arg Ala
    535                 540                 545

aca ata agc cct tta cac ttg gct gcc tat cat ggt cac cat caa gca      2154
Thr Ile Ser Pro Leu His Leu Ala Ala Tyr His Gly His His Gln Ala
550                 555                 560                 565

ctg gaa gtg ttg gta cag tct ttg tta gat ctt gat gtc aga aat agt      2202
Leu Glu Val Leu Val Gln Ser Leu Leu Asp Leu Asp Val Arg Asn Ser
                570                 575                 580

agt gga aga aca ccc cta gat ctt gca gct ttt aag ggc cat gtt gaa      2250
```

```
Ser Gly Arg Thr Pro Leu Asp Leu Ala Ala Phe Lys Gly His Val Glu
            585                 590                 595

tgt gtg gat gta ctc att aat cag gga gcc tca atc tta gta aaa gat      2298
Cys Val Asp Val Leu Ile Asn Gln Gly Ala Ser Ile Leu Val Lys Asp
        600                 605                 610

tac att ttg aag agg aca cct att cat gca gca gca aca aat ggt cat      2346
Tyr Ile Leu Lys Arg Thr Pro Ile His Ala Ala Ala Thr Asn Gly His
    615                 620                 625

tca gaa tgc tta cgg cta tta ata gga aat gca gaa cca cag aat gca      2394
Ser Glu Cys Leu Arg Leu Leu Ile Gly Asn Ala Glu Pro Gln Asn Ala
630                 635                 640                 645

gtg gat att caa gat gga aat gga cag acg cct ctg atg cta tct gtt      2442
Val Asp Ile Gln Asp Gly Asn Gly Gln Thr Pro Leu Met Leu Ser Val
                650                 655                 660

ctc aac ggg cac aca gac tgt gtt tac tca ttg ctg aac aaa gga gca      2490
Leu Asn Gly His Thr Asp Cys Val Tyr Ser Leu Leu Asn Lys Gly Ala
            665                 670                 675

aat gta gat gcc aaa gat aag tgg gga agg aca gcg ttg cat aga ggg      2538
Asn Val Asp Ala Lys Asp Lys Trp Gly Arg Thr Ala Leu His Arg Gly
        680                 685                 690

gca gtt aca ggc cat gaa gaa tgt gta gat gca tta ctt caa cat ggt      2586
Ala Val Thr Gly His Glu Glu Cys Val Asp Ala Leu Leu Gln His Gly
    695                 700                 705

gct aag tgc tta ctt cgg gat agc agg ggc cgg acg cct ata cac ctg      2634
Ala Lys Cys Leu Leu Arg Asp Ser Arg Gly Arg Thr Pro Ile His Leu
710                 715                 720                 725

tct gct gcc tgt gga cac att ggt gtt ctt gga gcc ctt ttg cag tca      2682
Ser Ala Ala Cys Gly His Ile Gly Val Leu Gly Ala Leu Leu Gln Ser
                730                 735                 740

gca gca tct atg gat gca aat cca gcc aca gca gac aat cat gga tat      2730
Ala Ala Ser Met Asp Ala Asn Pro Ala Thr Ala Asp Asn His Gly Tyr
            745                 750                 755

acg gca ctt cac tgg gct tgc tac aat ggt cac gag aca tgt gta gaa      2778
Thr Ala Leu His Trp Ala Cys Tyr Asn Gly His Glu Thr Cys Val Glu
        760                 765                 770

ctg ctt tta gaa cag gaa gtt ttc cag aaa acg gaa gga aat gct ttt      2826
Leu Leu Leu Glu Gln Glu Val Phe Gln Lys Thr Glu Gly Asn Ala Phe
    775                 780                 785

agt cca ttg cat tgt gcc gtg ata aat gac aac gaa ggt gct gct gag      2874
Ser Pro Leu His Cys Ala Val Ile Asn Asp Asn Glu Gly Ala Ala Glu
790                 795                 800                 805

atg tta att gat aca tta ggt gcc agc att gtg aac gcc aca gat tca      2922
Met Leu Ile Asp Thr Leu Gly Ala Ser Ile Val Asn Ala Thr Asp Ser
```

```
                  810                 815                 820

aaa gga aga act cct ctc cat gca gcc gcc ttc aca gac cat gta gag     2970
Lys Gly Arg Thr Pro Leu His Ala Ala Ala Phe Thr Asp His Val Glu
            825                 830                 835

tgt tta cag ctg ctg ctc agc cat aat gct caa gtc aat tct gtg gac     3018
Cys Leu Gln Leu Leu Leu Ser His Asn Ala Gln Val Asn Ser Val Asp
        840                 845                 850

tct aca ggg aaa aca cct ctt atg atg gct gca gaa aat gga caa aca     3066
Ser Thr Gly Lys Thr Pro Leu Met Met Ala Ala Glu Asn Gly Gln Thr
    855                 860                 865

aat aca gtt gag atg ctg gtt agc agt gct agt gca gaa ctg act tta     3114
Asn Thr Val Glu Met Leu Val Ser Ser Ala Ser Ala Glu Leu Thr Leu
870                 875                 880                 885

caa gat aac agt aaa aat act gcc ctc cat ttg gct tgt agc aag ggt     3162
Gln Asp Asn Ser Lys Asn Thr Ala Leu His Leu Ala Cys Ser Lys Gly
                890                 895                 900

cat gaa act agt gcc ttg tta ata ctg gaa aag ata aca gat aga aac     3210
His Glu Thr Ser Ala Leu Leu Ile Leu Glu Lys Ile Thr Asp Arg Asn
            905                 910                 915

ctc atc aat gca acc aac gca gcc ttg caa aca cct ctg cat gtt gct     3258
Leu Ile Asn Ala Thr Asn Ala Ala Leu Gln Thr Pro Leu His Val Ala
        920                 925                 930

gcc cga aat ggg cta aca atg gtg gtt cag gaa ctt ttg gga aaa gga     3306
Ala Arg Asn Gly Leu Thr Met Val Val Gln Glu Leu Leu Gly Lys Gly
    935                 940                 945

gca agt gtg ctt gca gta gat gaa aat ggc tat acc cca gct ttg gcc     3354
Ala Ser Val Leu Ala Val Asp Glu Asn Gly Tyr Thr Pro Ala Leu Ala
950                 955                 960                 965

tgt gct ccc aat aag gat gtg gct gat tgc ctg gct ctc att ttg gcc     3402
Cys Ala Pro Asn Lys Asp Val Ala Asp Cys Leu Ala Leu Ile Leu Ala
                970                 975                 980

acc atg atg cct gtc tca tca agt agt cct tta tca tcc tta aca ttc     3450
Thr Met Met Pro Val Ser Ser Ser Ser Pro Leu Ser Ser Leu Thr Phe
            985                 990                 995

aat gcc att aac cgt tat acc aac acc tca aaa aca gtc agc ttt gaa     3498
Asn Ala Ile Asn Arg Tyr Thr Asn Thr Ser Lys Thr Val Ser Phe Glu
       1000                1005                1010

gct ttg ccc atc atg agg aat gaa cct agc tcc tat tgc agt ttc aat     3546
Ala Leu Pro Ile Met Arg Asn Glu Pro Ser Ser Tyr Cys Ser Phe Asn
   1015                1020                1025

aac att gga ggg gaa cag gag tac tta tac act gac gtg gat gag ctc     3594
Asn Ile Gly Gly Glu Gln Glu Tyr Leu Tyr Thr Asp Val Asp Glu Leu
1030                1035                1040                1045
```

```
aac gac tcc gat tct gag acc tac tga gaggc tgaggaggag ggagttctca     3646
Asn Asp Ser Asp Ser Glu Thr Tyr  *
               1050

cagtaaagct tcaaactgtg ctttttcagg aaacaggcac tttgatattc acgtagaaat   3706

tcaacctaag agggacagtt ccgttgaagc cgt                               3739
```

<210> 8
<211> 2162
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (387)..(1748)

<400> 8

```
cgtactgaag tggtctgcat gagcgaccgg tccggaattc ccgggtcgac gatttcgttc     60

cggttgcact cttcctatag cccagagggc gagagggcct gtggcctggg ggaaggagga    120

cgaggttctg cctggatccc agcagtagga cgctgtgcca tttgggaaca aaggaatagt    180

ctgcctggaa tccctgcaga tcttggggcc ggaggccagt ccaacccttg gagcaggaag    240

aaacgcaaag ttgtcaagaa ccaagtcgag ctgcctcaga gccggcccgc agtagctgca    300

gactccgccc gcgacgtgtg cgcgcttctc tgggccagag cgagcctgtt ttgtgctcgg    360

gttaagagat ttgtcccagc tatacc atg ggc cgc act cgg gaa gct ggc tgc      413
                              Met Gly Arg Thr Arg Glu Ala Gly Cys
                                1               5

gtg gcc gct ggt gtg gtt atc ggg gct ggt gcc tgc tac tgt gta tac      461
Val Ala Ala Gly Val Val Ile Gly Ala Gly Ala Cys Tyr Cys Val Tyr
 10                  15                  20                  25

aga ctg gct tgg gga aga gac gag aac gag aaa atc tgg gac gaa gac      509
Arg Leu Ala Trp Gly Arg Asp Glu Asn Glu Lys Ile Trp Asp Glu Asp
                 30                  35                  40

gag gag tct acg gac acc tca gag att ggg gtt gag act gtg aaa gga      557
Glu Glu Ser Thr Asp Thr Ser Glu Ile Gly Val Glu Thr Val Lys Gly
             45                  50                  55

gct aaa act aac gct ggg gca ggg tct ggg gcc aaa ctt cag ggt gat      605
Ala Lys Thr Asn Ala Gly Ala Gly Ser Gly Ala Lys Leu Gln Gly Asp
         60                  65                  70

tca gag gtc aag cct gag gtg agt ttg gga ctc gag gat tgt ccg ggt      653
Ser Glu Val Lys Pro Glu Val Ser Leu Gly Leu Glu Asp Cys Pro Gly
     75                  80                  85
```

```
gta aaa gag aag gcc cat tca gga tcc cac agc gga ggt ggt cta gag      701
Val Lys Glu Lys Ala His Ser Gly Ser His Ser Gly Gly Gly Leu Glu
 90                  95                  100                 105

gcc aag gcc aag gcc ctt ttc aac acg ctg aag gaa cag gca agt gca      749
Ala Lys Ala Lys Ala Leu Phe Asn Thr Leu Lys Glu Gln Ala Ser Ala
                110                 115                 120

aag gca ggc aaa ggg gct agg gtg ggt acc atc tct ggg aac agg acc      797
Lys Ala Gly Lys Gly Ala Arg Val Gly Thr Ile Ser Gly Asn Arg Thr
            125                 130                 135

ctt gca ccg agt tta ccc tgc cca gga ggc agg ggt gga ggc tgc cac      845
Leu Ala Pro Ser Leu Pro Cys Pro Gly Gly Arg Gly Gly Gly Cys His
        140                 145                 150

ccc acc agg agt gga tct agg gcc ggg ggc agg gca agt gga aaa tcc      893
Pro Thr Arg Ser Gly Ser Arg Ala Gly Gly Arg Ala Ser Gly Lys Ser
    155                 160                 165

aag gga aag gcc cga agt aag agc acc agg gct cca gct aca aca tgg      941
Lys Gly Lys Ala Arg Ser Lys Ser Thr Arg Ala Pro Ala Thr Thr Trp
170                 175                 180                 185

cct gtc cgg aga ggc aag ttc aac ttt cct tat aaa att gat gat att      989
Pro Val Arg Arg Gly Lys Phe Asn Phe Pro Tyr Lys Ile Asp Asp Ile
                190                 195                 200

ctg agt gct ccc gac ctc caa aag gtc ctc aac atc ctg gag cga aca     1037
Leu Ser Ala Pro Asp Leu Gln Lys Val Leu Asn Ile Leu Glu Arg Thr
            205                 210                 215

aat gat cct ttt att caa gaa gta gcc ttg gtc act ctg ggt aac aat     1085
Asn Asp Pro Phe Ile Gln Glu Val Ala Leu Val Thr Leu Gly Asn Asn
        220                 225                 230

gca gca tat tca ttt aac cag aat gcc ata cgt gaa ttg ggt ggt gtc     1133
Ala Ala Tyr Ser Phe Asn Gln Asn Ala Ile Arg Glu Leu Gly Gly Val
    235                 240                 245

cca att att gca aaa ctg ata aaa aca aaa gac ccc ata att agg gaa     1181
Pro Ile Ile Ala Lys Leu Ile Lys Thr Lys Asp Pro Ile Ile Arg Glu
250                 255                 260                 265

aag act tac aat gcc ctt aat aac ttg agt gtg aac gca gaa aat cag     1229
Lys Thr Tyr Asn Ala Leu Asn Asn Leu Ser Val Asn Ala Glu Asn Gln
                270                 275                 280

ggc aag att aag acg tac atc agt caa gtg tgt gat gac acc atg gtc     1277
Gly Lys Ile Lys Thr Tyr Ile Ser Gln Val Cys Asp Asp Thr Met Val
            285                 290                 295

tgt cgc ttg gac tca gct gtg cag atg gct ggg cta aga ctg tta acc     1325
Cys Arg Leu Asp Ser Ala Val Gln Met Ala Gly Leu Arg Leu Leu Thr
        300                 305                 310
```

```
aac atg act gtg act aat cat tac caa cat ttg ctt tcc tat tct ttt     1373
Asn Met Thr Val Thr Asn His Tyr Gln His Leu Leu Ser Tyr Ser Phe
    315                 320                 325

cca gac ttt ttt gct ttg tta ttc ctg gga aat cac ttc acc aag ata     1421
Pro Asp Phe Phe Ala Leu Leu Phe Leu Gly Asn His Phe Thr Lys Ile
330                 335                 340                 345

cag att atg aaa cta att ata aac ttt act gaa aat cca gcc atg aca     1469
Gln Ile Met Lys Leu Ile Ile Asn Phe Thr Glu Asn Pro Ala Met Thr
                350                 355                 360

aga gag ctg gtc agt tgt aaa gta cca tca gaa ttg att tcc ctc ttt     1517
Arg Glu Leu Val Ser Cys Lys Val Pro Ser Glu Leu Ile Ser Leu Phe
            365                 370                 375

aat aaa gaa tgg gat aga gag att ctt ctt aat atc ctt acc cta ttt     1565
Asn Lys Glu Trp Asp Arg Glu Ile Leu Leu Asn Ile Leu Thr Leu Phe
        380                 385                 390

gag aat ata aat gac aac ata aaa aat gaa ggg ctc gca tca tcc agg     1613
Glu Asn Ile Asn Asp Asn Ile Lys Asn Glu Gly Leu Ala Ser Ser Arg
    395                 400                 405

aaa gaa ttc agc aga agt tca ctt ttt ttc tta ttc aaa gag tct gga     1661
Lys Glu Phe Ser Arg Ser Ser Leu Phe Phe Leu Phe Lys Glu Ser Gly
410                 415                 420                 425

gtt tgt gtt aag aaa atc aaa gca cta gca aat cac aat gat ctg gtg     1709
Val Cys Val Lys Lys Ile Lys Ala Leu Ala Asn His Asn Asp Leu Val
                430                 435                 440

gtg aaa gta aaa gtc ctg aaa gta tta acc aaa ctc taa tttggagtct      1758
Val Lys Val Lys Val Leu Lys Val Leu Thr Lys Leu  *
            445                 450

gtcccaaaca atattgagat atttgcagtt ggtacgatgt gatttgtaaa ttctttgttt   1818

ttcattgtgc gtatatggta aagagatctt ttcagctgct attttggaat aatgactatc   1878

atatatcata acagtgactg atgttggttg taatggttgg gtttaggatg aaccatttta   1938

aggatgccaa atgaaatatt agtatttgta cacagaaaga atttattgat ttgatcttat   1998

tacctagatt gagatttttt aatctttcct ctacctaaac tgacaatgaa ttggttatac   2058

atcatgcata agctacactt ttatattagt ttatatttgt tattctaaga cttgtgtttc   2118

atcaataaag ttgtgtttta agcagcagaa aaaaaaaaaa aaaa                    2162
```

<210> 9
<211> 1614
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (202)..(969)

<400> 9

```
aagctggtac gcctgcaggt accggtccgg aattcccggg tcgacgattt cgtcaaagag     60

acaaactcca ttttcttatg aatggaaagt gaaaacccct gttccgctta aattgggttc    120

cttcctgtcc tgagaaacat agagaccccc aaaagggaag cagaggagag aaagtcccac    180

acccagaccc cgcgagaaga g  atg acc atg acc acc atg cca gaa agt ctc     231
                          Met Thr Met Thr Thr Met Pro Glu Ser Leu
                            1               5                  10

aac agc ccc gtg tcg ggc aag gcg gtg ttt atg gag ttt ggg ccg ccc      279
Asn Ser Pro Val Ser Gly Lys Ala Val Phe Met Glu Phe Gly Pro Pro
                 15                  20                  25

aac cag caa atg tct cct tcc ccc atg tcc cac ggg cac tac tcc atg      327
Asn Gln Gln Met Ser Pro Ser Pro Met Ser His Gly His Tyr Ser Met
             30                  35                  40

cac tgt tta cac tcg gcg ggc cat tcg cag ccc gac ggc gcc tac agc      375
His Cys Leu His Ser Ala Gly His Ser Gln Pro Asp Gly Ala Tyr Ser
         45                  50                  55

tca gcc tcg tcc ttc tcc cga ccg ctg ggc tac ccc tac gtc aac tcg      423
Ser Ala Ser Ser Phe Ser Arg Pro Leu Gly Tyr Pro Tyr Val Asn Ser
     60                  65                  70

gtc agc agc cac gca tcc agc ccc tac atc agt tcg gtg cag tcc tac      471
Val Ser Ser His Ala Ser Ser Pro Tyr Ile Ser Ser Val Gln Ser Tyr
 75                  80                  85                  90

ccg ggc agc gcc agc ctc gcc cag agc cgc ctg gag gac cca ggg gcg      519
Pro Gly Ser Ala Ser Leu Ala Gln Ser Arg Leu Glu Asp Pro Gly Ala
                 95                 100                 105

gac tcg gag aag agc acg gtg gtg gaa ggc ggt gaa gtg cgc ttc aat      567
Asp Ser Glu Lys Ser Thr Val Val Glu Gly Gly Glu Val Arg Phe Asn
            110                 115                 120

ggc aag gga aaa aag atc cgt aaa ccc agg acg att tat tcc agt ttg      615
Gly Lys Gly Lys Lys Ile Arg Lys Pro Arg Thr Ile Tyr Ser Ser Leu
        125                 130                 135

cag ttg cag gct ttg aac cgg agg ttc cag caa act cag tac cta gct      663
Gln Leu Gln Ala Leu Asn Arg Arg Phe Gln Gln Thr Gln Tyr Leu Ala
    140                 145                 150

ctg ccg gag agg gcg gag ctc gcg gcc tct ttg gga ctc aca cag act      711
Leu Pro Glu Arg Ala Glu Leu Ala Ala Ser Leu Gly Leu Thr Gln Thr
155                 160                 165                 170

cag gtc aag atc tgg ttc caa aac aag cga tcc aag ttc aag aag ctg      759
```

```
Gln Val Lys Ile Trp Phe Gln Asn Lys Arg Ser Lys Phe Lys Lys Leu
                175                 180                 185

atg aag cag ggt ggg gcg gct ctg gag ggt agt gcg ttg gcc aac ggg      807
Met Lys Gln Gly Gly Ala Ala Leu Glu Gly Ser Ala Leu Ala Asn Gly
            190                 195                 200

cgg gcc ctg tct gct ggc tcc cca ccc gtg ccg ccc ggc tgg aac cct      855
Arg Ala Leu Ser Ala Gly Ser Pro Pro Val Pro Pro Gly Trp Asn Pro
        205                 210                 215

aac tct tca tcc ggg aag ggc tca gga gga aac gcg ggc tcc tat atc      903
Asn Ser Ser Ser Gly Lys Gly Ser Gly Gly Asn Ala Gly Ser Tyr Ile
    220                 225                 230

ccc agc tac aca tcg tgg tac cct tca gcg cac caa gaa gct atg cag      951
Pro Ser Tyr Thr Ser Trp Tyr Pro Ser Ala His Gln Glu Ala Met Gln
235                 240                 245                 250

caa ccc caa ctt atg tga ggttgc ccgcccgtct ccttcttgtc tccccggccc     1005
Gln Pro Gln Leu Met  *
                255

aggtccctcc cgcctccagg tccatccatc ccgtccggaa aagaaggacc cagagggaag   1065

aaggaacagt ggaggcggga cgccctccat ctcctcggag ccccgcgagg tccggcccag   1125

caacttcccg gcatccgcgc tctagcctga accctggcct gggccgagca gtggcagcag   1185

agagtggcct cggagggaag ccactgccac ctgagacagc ccaagcagca agataaaccc   1245

gctccacccg acccgccgac cttcagcttt gtgggactat caggaaaaaa caaaacaaaa   1305

acaaaatgta gaaaaagcaa aagctctttt ctgtcctgtc agtctcctgt ctccttttgc   1365

tctgtctgtg cgctggtaaa gtccaggtcc tcatccgtcc gctgtcctca ttctgcggcc   1425

tcagcaaaaa gccacaaggt ctgagcggcc cgggtcctgc cgggctgacc atctccggat   1485

cctgggacac tctgcctgac catctgtgta gctggtgtgg gaatctgggg gcattggagg   1545

gaggggttt acgaaatcgt cgacagcgaa gtcgacccgg gaattccggg ccggtacgtg   1605

caggcgtac                                                          1614
```

<210> 10
<211> 4351
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (154)..(3240)

<400> 10

```
ccctctaatc tgatcactat agggaatttg gccctcgagc cgaagaattc ggcacgaggt     60

ctggctggga gcagaaggca gcctcggtct ctgggcggcg gcggcggccc actctgccct    120

ggccgcgctg tgtggtgacc gcaggcccga gac atg agg gcg gcc cgt gct ctg     174
                                        Met Arg Ala Ala Arg Ala Leu
                                          1               5

ctg ccc ctg ctg ctg cag gcc tgc tgg aca gcc gcg cag gat gag ccg      222
Leu Pro Leu Leu Leu Gln Ala Cys Trp Thr Ala Ala Gln Asp Glu Pro
         10                  15                  20

gag acc ccg agg gcc gtg gcc ttc cag gac tgc ccc gtg gac ctg ttc      270
Glu Thr Pro Arg Ala Val Ala Phe Gln Asp Cys Pro Val Asp Leu Phe
     25                  30                  35

ttt gtg ctg gac acc tct gag agc gtg gcc ctg agg ctg aag ccc tac      318
Phe Val Leu Asp Thr Ser Glu Ser Val Ala Leu Arg Leu Lys Pro Tyr
 40                  45                  50                  55

ggg gcc ctc gtg gac aaa gtc aag tcc ttc acc aag cgc ttc atc gac      366
Gly Ala Leu Val Asp Lys Val Lys Ser Phe Thr Lys Arg Phe Ile Asp
                 60                  65                  70

aac ctg agg gac agg tac tac cgc tgt gac cga aac ctg gtg tgg aac      414
Asn Leu Arg Asp Arg Tyr Tyr Arg Cys Asp Arg Asn Leu Val Trp Asn
             75                  80                  85

gca ggc gcg ctg cac tac agt gac gag gtg gag atc atc caa ggc ctc      462
Ala Gly Ala Leu His Tyr Ser Asp Glu Val Glu Ile Ile Gln Gly Leu
         90                  95                  100

acg cgc atg cct ggc ggc cgc gac gca ctc aaa agc agc gtg gac gcg      510
Thr Arg Met Pro Gly Gly Arg Asp Ala Leu Lys Ser Ser Val Asp Ala
    105                 110                 115

gtc aag tac ttt ggg aag ggc acc tac acc gac tgc gct atc aag aag      558
Val Lys Tyr Phe Gly Lys Gly Thr Tyr Thr Asp Cys Ala Ile Lys Lys
120                 125                 130                 135

ggg ctg gag cag ctc ctc gtg ggg ggc tcc cac ctg aag gag aat aag      606
Gly Leu Glu Gln Leu Leu Val Gly Gly Ser His Leu Lys Glu Asn Lys
                140                 145                 150

tac ctg att gtg gtg acc gac ggg cac ccc ctg gag ggc tac aag gaa      654
Tyr Leu Ile Val Val Thr Asp Gly His Pro Leu Glu Gly Tyr Lys Glu
            155                 160                 165

ccc tgt ggg ggg ctg gag gat gct gtg aac gag gcc aag cac ctg ggc      702
Pro Cys Gly Gly Leu Glu Asp Ala Val Asn Glu Ala Lys His Leu Gly
        170                 175                 180

gtc aaa gtc ttc tcg gtg gcc atc aca ccc gac cac ctg gag ccg cgt      750
Val Lys Val Phe Ser Val Ala Ile Thr Pro Asp His Leu Glu Pro Arg
    185                 190                 195
```

```
ctg agc atc atc gcc acg gac cac acg tac cgg cgc aac ttc acg gcg      798
Leu Ser Ile Ile Ala Thr Asp His Thr Tyr Arg Arg Asn Phe Thr Ala
200                 205                 210                 215

gct gac tgg ggc cag agc cgc gac gca gag gag gcc atc agc cag acc      846
Ala Asp Trp Gly Gln Ser Arg Asp Ala Glu Glu Ala Ile Ser Gln Thr
                220                 225                 230

atc gac acc atc gtg gac atg atc aaa aat aac gtg gag caa gtg tgc      894
Ile Asp Thr Ile Val Asp Met Ile Lys Asn Asn Val Glu Gln Val Cys
            235                 240                 245

tgc tcc ttc gaa tgc cag cct gca aga gga cct ccg ggg ctc cgg ggc      942
Cys Ser Phe Glu Cys Gln Pro Ala Arg Gly Pro Pro Gly Leu Arg Gly
        250                 255                 260

gac ccc ggc ttt gag gga gaa cga ggc aag ccg ggg ctc cca gga gag      990
Asp Pro Gly Phe Glu Gly Glu Arg Gly Lys Pro Gly Leu Pro Gly Glu
    265                 270                 275

aag gga gaa gcc gga gat cct gga aga ccc ggg gac ctc gga cct gtt     1038
Lys Gly Glu Ala Gly Asp Pro Gly Arg Pro Gly Asp Leu Gly Pro Val
280                 285                 290                 295

ggg tac cag gga atg aag gga gaa aaa ggg agc cgt ggg gag aag ggc     1086
Gly Tyr Gln Gly Met Lys Gly Glu Lys Gly Ser Arg Gly Glu Lys Gly
                300                 305                 310

tcc agg gga ccc aag ggc tac aag gga gag aag ggc aag cgt ggc atc     1134
Ser Arg Gly Pro Lys Gly Tyr Lys Gly Glu Lys Gly Lys Arg Gly Ile
            315                 320                 325

gac ggg gtg gac ggc gtg aag ggg gag atg ggg tac cca ggc ctg cca     1182
Asp Gly Val Asp Gly Val Lys Gly Glu Met Gly Tyr Pro Gly Leu Pro
        330                 335                 340

ggc tgc aag ggc tcg ccc ggg ttt gac ggc att caa gga ccc cct ggc     1230
Gly Cys Lys Gly Ser Pro Gly Phe Asp Gly Ile Gln Gly Pro Pro Gly
    345                 350                 355

ccc aag gga gac ccc ggc gcc ttt gga ctg aaa gga gaa aag ggc gag     1278
Pro Lys Gly Asp Pro Gly Ala Phe Gly Leu Lys Gly Glu Lys Gly Glu
360                 365                 370                 375

cct gga gct gac ggg gag gcg ggg aga cca ggg agc tcg gga cca tct     1326
Pro Gly Ala Asp Gly Glu Ala Gly Arg Pro Gly Ser Ser Gly Pro Ser
                380                 385                 390

gga gac gag ggc cag ccg gga gag cct ggg ccc ccc gga gag aaa gga     1374
Gly Asp Glu Gly Gln Pro Gly Glu Pro Gly Pro Pro Gly Glu Lys Gly
            395                 400                 405

gag gcg ggc gac gag ggg aac cca gga cct gac ggt gcc ccc ggg gag     1422
Glu Ala Gly Asp Glu Gly Asn Pro Gly Pro Asp Gly Ala Pro Gly Glu
        410                 415                 420

cgg ggt ggc cct gga gag aga gga cca cgg ggg acc cca ggc acg cgg     1470
```

```
Arg Gly Gly Pro Gly Glu Arg Gly Pro Arg Gly Thr Pro Gly Thr Arg
    425                 430                 435

gga cca aga gga gac cct ggt gaa gct ggc ccg cag ggt gat cag gga      1518
Gly Pro Arg Gly Asp Pro Gly Glu Ala Gly Pro Gln Gly Asp Gln Gly
440                 445                 450                 455

aga gaa ggc ccc gtt ggt gtc cct gga gac ccg ggc gag gct ggc cct      1566
Arg Glu Gly Pro Val Gly Val Pro Gly Asp Pro Gly Glu Ala Gly Pro
                460                 465                 470

atc gga cct aaa ggc tac cga ggc gat gag ggt ccc cca ggg tcc gag      1614
Ile Gly Pro Lys Gly Tyr Arg Gly Asp Glu Gly Pro Pro Gly Ser Glu
            475                 480                 485

ggt gcc aga gga gcc cca gga cct gcc gga ccc cct gga gac ccg ggg      1662
Gly Ala Arg Gly Ala Pro Gly Pro Ala Gly Pro Pro Gly Asp Pro Gly
        490                 495                 500

ctg atg ggt gaa agg gga gaa gac ggc ccc gct gga aat ggc acc gag      1710
Leu Met Gly Glu Arg Gly Glu Asp Gly Pro Ala Gly Asn Gly Thr Glu
    505                 510                 515

ggc ttc ccc ggc ttc ccc ggg tat ccg ggc aac agg ggc gct ccc ggg      1758
Gly Phe Pro Gly Phe Pro Gly Tyr Pro Gly Asn Arg Gly Ala Pro Gly
520                 525                 530                 535

ata aac ggc acg aag ggc tac ccc ggc ctc aag ggg gac gag gga gaa      1806
Ile Asn Gly Thr Lys Gly Tyr Pro Gly Leu Lys Gly Asp Glu Gly Glu
                540                 545                 550

gcc ggg gac ccc gga gac gat aac aac gac att gca ccc cga gga gtc      1854
Ala Gly Asp Pro Gly Asp Asp Asn Asn Asp Ile Ala Pro Arg Gly Val
            555                 560                 565

aaa gga gca aag ggg tac cgg ggt ccc gag ggc ccc cag gga ccc cca      1902
Lys Gly Ala Lys Gly Tyr Arg Gly Pro Glu Gly Pro Gln Gly Pro Pro
        570                 575                 580

gga cac caa gga ccg cct ggg ccg gac gaa tgc gag att ttg gac atc      1950
Gly His Gln Gly Pro Pro Gly Pro Asp Glu Cys Glu Ile Leu Asp Ile
    585                 590                 595

atc atg aaa atg tgc tct tgc tgt gaa tgc aag tgc ggc ccc atc gac      1998
Ile Met Lys Met Cys Ser Cys Cys Glu Cys Lys Cys Gly Pro Ile Asp
600                 605                 610                 615

ctc ctg ttc gtg ctg gac agc tca gag agc att ggc ctg cag aac ttc      2046
Leu Leu Phe Val Leu Asp Ser Ser Glu Ser Ile Gly Leu Gln Asn Phe
                620                 625                 630

gag att gcc aag gac ttc gtc gtc aag gtc atc gac cgg ctg agc cgg      2094
Glu Ile Ala Lys Asp Phe Val Val Lys Val Ile Asp Arg Leu Ser Arg
            635                 640                 645

gac gag ctg gtc aag ttc gag cca ggg cag tcg tac gcg ggt gtg gtg      2142
Asp Glu Leu Val Lys Phe Glu Pro Gly Gln Ser Tyr Ala Gly Val Val
```

```
        650                 655                 660

cag tac agc cac agc cag atg cag gag cac gtg agc ctg cgc agc ccc     2190
Gln Tyr Ser His Ser Gln Met Gln Glu His Val Ser Leu Arg Ser Pro
    665                 670                 675

agc atc cgg aac gtg cag gag ctc aag gaa gcc atc aag agc ctg cag     2238
Ser Ile Arg Asn Val Gln Glu Leu Lys Glu Ala Ile Lys Ser Leu Gln
680                 685                 690                 695

tgg atg gcg ggc ggc acc ttc acg ggg gag gcc ctg cag tac acg cgg     2286
Trp Met Ala Gly Gly Thr Phe Thr Gly Glu Ala Leu Gln Tyr Thr Arg
                700                 705                 710

gac cag ctg ctg ccg ccc agc ccg aac aac cgc atc gcc ctg gtc atc     2334
Asp Gln Leu Leu Pro Pro Ser Pro Asn Asn Arg Ile Ala Leu Val Ile
            715                 720                 725

act gac ggg cgc tca gac act cag agg gac acc aca ccg ctc aac gtg     2382
Thr Asp Gly Arg Ser Asp Thr Gln Arg Asp Thr Thr Pro Leu Asn Val
        730                 735                 740

ctc tgc agc ccc ggc atc cag gtg gtc tcc gtg ggc atc aaa gac gtg     2430
Leu Cys Ser Pro Gly Ile Gln Val Val Ser Val Gly Ile Lys Asp Val
    745                 750                 755

ttt gac ttc atc cca ggc tca gac cag ctc aat gtc att tct tgc caa     2478
Phe Asp Phe Ile Pro Gly Ser Asp Gln Leu Asn Val Ile Ser Cys Gln
760                 765                 770                 775

ggc ctg gca cca tcc cag ggc cgg ccc ggc ctc tcg ctg gtc aag gag     2526
Gly Leu Ala Pro Ser Gln Gly Arg Pro Gly Leu Ser Leu Val Lys Glu
                780                 785                 790

aac tat gca gag ctg ctg gag gat gcc ttc ctg aag aat gtc acc gcc     2574
Asn Tyr Ala Glu Leu Leu Glu Asp Ala Phe Leu Lys Asn Val Thr Ala
            795                 800                 805

cag atc tgc ata gac aag aag tgt cca gat tac acc tgc ccc atc acg     2622
Gln Ile Cys Ile Asp Lys Lys Cys Pro Asp Tyr Thr Cys Pro Ile Thr
        810                 815                 820

ttc tcc tcc ccg gct gac atc acc atc ctg ctg gac ggc tcc gcc agc     2670
Phe Ser Ser Pro Ala Asp Ile Thr Ile Leu Leu Asp Gly Ser Ala Ser
    825                 830                 835

gtg ggc agc cac aac ttt gac acc acc aag cgc ttc gcc aag cgc ctg     2718
Val Gly Ser His Asn Phe Asp Thr Thr Lys Arg Phe Ala Lys Arg Leu
840                 845                 850                 855

gcc gag cgc ttc ctc aca gcg ggc agg acg gac ccc gcc cac gac gtg     2766
Ala Glu Arg Phe Leu Thr Ala Gly Arg Thr Asp Pro Ala His Asp Val
                860                 865                 870

cgg gtg gcg gtg gtg cag tac agc ggc acg ggc cag cag cgc cca gag     2814
Arg Val Ala Val Val Gln Tyr Ser Gly Thr Gly Gln Gln Arg Pro Glu
            875                 880                 885
```

```
cgg gcg tcg ctg cag ttc ctg cag aac tac acg gcc ctg gcc agt gcc     2862
Arg Ala Ser Leu Gln Phe Leu Gln Asn Tyr Thr Ala Leu Ala Ser Ala
        890                 895                 900

gtc gat gcc atg gac ttt atc aac gac gcc acc gac gtc aac gat gcc     2910
Val Asp Ala Met Asp Phe Ile Asn Asp Ala Thr Asp Val Asn Asp Ala
    905                 910                 915

ctg ggc tat gtg acc cgc ttc tac cgc gag gcc tcg tcc ggc gct gcc     2958
Leu Gly Tyr Val Thr Arg Phe Tyr Arg Glu Ala Ser Ser Gly Ala Ala
920                 925                 930                 935

aag aag agg ctg ctg ctc ttc tca gat ggc aac tcg cag ggc gcc acg     3006
Lys Lys Arg Leu Leu Leu Phe Ser Asp Gly Asn Ser Gln Gly Ala Thr
                940                 945                 950

ccc gct gcc atc gag aag gcc gtg cag gaa gcc cag cgg gca ggc atc     3054
Pro Ala Ala Ile Glu Lys Ala Val Gln Glu Ala Gln Arg Ala Gly Ile
            955                 960                 965

gag atc ttc gtg gtg gtc gtg ggc cgc cag gtg aat gag ccc cac atc     3102
Glu Ile Phe Val Val Val Val Gly Arg Gln Val Asn Glu Pro His Ile
        970                 975                 980

cgc gtc ctg gtc acc ggc aag acg gcc gag tac gac gtg gcc tac ggc     3150
Arg Val Leu Val Thr Gly Lys Thr Ala Glu Tyr Asp Val Ala Tyr Gly
    985                 990                 995

gag agc cac ctg ttc cgt gtc ccc agc tac cag gcc ctg ctc cgc ggt     3198
Glu Ser His Leu Phe Arg Val Pro Ser Tyr Gln Ala Leu Leu Arg Gly
1000                1005                1010                1015

gtc ttc cac cag aca gtc tcc agg aag gtg gcg ctg ggc tag cccaccc     3247
Val Phe His Gln Thr Val Ser Arg Lys Val Ala Leu Gly  *
                1020                1025

tgcacgccgg caccaaaccc tgtcctccca cccctcccca ctcatcacta aacagagaaa   3307

agcttggaaa gccaggacac aacgctgctg cctgctttgt gcagggtcct ccggggctca   3367

gccctgagtt ggcatcacct gcgcagggcc ctctggggct cagccctgag ctagtgtcac   3427

ctgcacaggg ccctctgggg ctcagccctg agctggcgtc acctgtgcag ggccctctgg   3487

ggctcagccc tgagctggcc tcacctgggt tccccacccc gggctctcct gccctgccct   3547

cctgcccgcc ctccctcctg cctgcgcagc tccttcccta ggcacctctg tgctgcgtcc   3607

caccagcctg agcaagacgc cctctcgggg cctgtgccgc actagcctcc ctctcctctg   3667

tccccatagc tggtttttcc caccaatcct cacctaacag ttactttaca attaaactca   3727

aagcaagctc ttctcctcag cttggggcag ccattggcct ctgtctcgtt ttgggaaacc   3787

aaggtcagga ggccgttgca gacataaatc tcggcgactc ggccccgtct cctgagggtc   3847
```

ctgctggtga ccggcctgga ccttggccct acagccctgg aggccgctgc tgaccagcac 3907

tgaccccgac ctcagagagt actcgcaggg gcgctggctg cactcaagac cctcgagatt 3967

aacggtgcta accccgtctg ctcctccctc ccgcagagac tggggcctgg actggacatg 4027

agagcccctt ggtgccacag agggctgtgt cttactagaa acaacgcaaa cctctccttc 4087

ctcagaatag tgatgtgttc gacgttttat caaaggcccc ctttctatgt tcatgttagt 4147

tttgctcctt ctgtgttttt ttctgaacca tatccatgtt gctgactttt ccaaataaag 4207

gttttcactc ctcaaaaaaa aaaaaaaaag ggcggccgct ctagagtatc cctcgagggg 4267

cccaagctta cgcgtaccca gctttcttgt acaaagtggt ccctatagtg agtcgtatat 4327

aagctaggga cgtttcagcg acag 4351

<210> 11
<211> 1622
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (864)..(1424)

<400> 11
taagcttgcg gccgctgaac tacagctctg cgcctgccca ggcggccgca cgctcagggg 60

cgtggcatgg gtgggtcgtg agttgggcgg ggcccacagg gcgtgcgcga cgcagcggcg 120

cggcgcgagg cgtaaggggc gtggcgccag tgggcgtggc gtggcgcagt gcgaagggac 180

gcggtgcgca tgcgcgtgag ggcttccacg ggtgggtggt atcgaggcct gtcgggtcag 240

ggcggttcgc gggtgctgtc agagctgggc cggggcccct aggcaggccc agacatgtcc 300

gtccttgtaa gttaaaagct tccatgggag ccttccttcc taatcaagat gcaaatagta 360

cggtattccg aacagacact aaaaatagct gtcatctcaa agaatccagt gcttgtgtca 420

cagtatgaga aagtagatgc tggggaacag cgtttaatga atgaggcatg caagccagcc 480

agtgatctct ttggaccttg cattctccat cagattggat cacctcccac cctgaggccc 540

cccaagactt tgaacagttc ttcagtcatc cttacagaaa gataccctct ccagacaaac 600

gcagtattta tatacggtcc attggatctc tatgaagcac cagaattatc agtgaagaat 660

atattaaatg gctcacgggc tactgtaaag catatttcta tcgcttgaga gtaaaactcc 720

tagaaccagt tcctgtttct gtaacaagat gttcctttag agtcaatgag aacacacaca 780

```
acctacaaat tcatgcaggg gacatcctga agttcttgaa aaagaagaaa cctgaagatg     840

ccttctgtgt tgtgggaata aca    atg att gat ctt tac cca aga gac tcg      890
                             Met Ile Asp Leu Tyr Pro Arg Asp Ser
                               1               5

tgg aat ttt gtc ttt gga cag gcc tct ttg aca gat ggt gtg ggg ata      938
Trp Asn Phe Val Phe Gly Gln Ala Ser Leu Thr Asp Gly Val Gly Ile
 10                  15                  20                  25

ttc agc ttt gcc agg tat ggc agt gat ttt tat agc atg cac tat aaa      986
Phe Ser Phe Ala Arg Tyr Gly Ser Asp Phe Tyr Ser Met His Tyr Lys
                 30                  35                  40

ggc aaa gtg aag aag ctc aag aaa aca tct tca agt gac tat tca att     1034
Gly Lys Val Lys Lys Leu Lys Lys Thr Ser Ser Ser Asp Tyr Ser Ile
             45                  50                  55

ttc gac aac tat tat att cca gaa ata act agt gtt tta cta ctt cga     1082
Phe Asp Asn Tyr Tyr Ile Pro Glu Ile Thr Ser Val Leu Leu Leu Arg
         60                  65                  70

tcc tgt aag act tta acc cat gag atc gga cac ata ttt gga ctg cga     1130
Ser Cys Lys Thr Leu Thr His Glu Ile Gly His Ile Phe Gly Leu Arg
     75                  80                  85

cac tgc cag tgg ctt gca tgc ctc atg caa ggc tcc aac cac ttg gaa     1178
His Cys Gln Trp Leu Ala Cys Leu Met Gln Gly Ser Asn His Leu Glu
 90                  95                 100                 105

gaa gct gac cgg cgc cct cta aac ctt tgc cct atc tgt ttg cac aag     1226
Glu Ala Asp Arg Arg Pro Leu Asn Leu Cys Pro Ile Cys Leu His Lys
                110                 115                 120

ttg cag tgt gct gtt ggc ttc agc att gta gaa aga tac aaa gca ctg     1274
Leu Gln Cys Ala Val Gly Phe Ser Ile Val Glu Arg Tyr Lys Ala Leu
            125                 130                 135

gtg agg tgg att gat gat gaa tct tct gac aca cct gga gca act cca     1322
Val Arg Trp Ile Asp Asp Glu Ser Ser Asp Thr Pro Gly Ala Thr Pro
        140                 145                 150

gaa cac agt cac gag gat aat ggg aat tta ccg aaa ccc gtg gaa gcc     1370
Glu His Ser His Glu Asp Asn Gly Asn Leu Pro Lys Pro Val Glu Ala
    155                 160                 165

ttt aag gaa tgg aaa gag tgg ata ata aaa tgc ctg gct gtt ctc caa     1418
Phe Lys Glu Trp Lys Glu Trp Ile Ile Lys Cys Leu Ala Val Leu Gln
170                 175                 180                 185

aaa tga ggaccttcaa ataggagtga ttgaaataaa taactacttg catgttatgc      1474
Lys  *

tttcatttgg gtggaatact tcattggaat aaactactga tcttgtgctg tgtcaaagta   1534

acagactaga accttctttc aagtacctga attgaaatga aactcatttt gaataataaa   1594
```

aactctagaa actcttaaaa aaaaaaaa 1622

<210> 12
<211> 882
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (147)..(284)

<400> 12
gagccgagat cacgccactg cactccagcc tgggcaacaa gagcaaaact ctgtctcaat 60

aaaaaaaaag gaaaacccac agccaacatt atcactagtg gtaacaggga aaatgtcccc 120

tttgaccccc accccaagta cagaac atg caa gga cgc ctg ttc tca ctg cat 173
Met Gln Gly Arg Leu Phe Ser Leu His
1 5

gtc ttc tgt gtt gta ctg aag ctc cta gct ggt gca gtg agg caa gac 221
Val Phe Cys Val Val Leu Lys Leu Leu Ala Gly Ala Val Arg Gln Asp
10 15 20 25

aaa gaa agg aaa tat agg tgc act ggg aag gaa gaa gaa aca ctg cct 269
Lys Glu Arg Lys Tyr Arg Cys Thr Gly Lys Glu Glu Glu Thr Leu Pro
30 35 40

tta ttc ttt agg tga catgattgtg tgcttttaaa ataataaagg aatcaacaga 324
Leu Phe Phe Arg *
45

aaagttgctt aacctaatga atgagtttat caaagtcaca agatacaagg tcagtataca 384

aaaatcagtt ggatttctac atggtagaaa caactgtaca tggaaaaatg tttaatagtg 444

taagatatgt acattggaaa ctatgaaaga gtgtaaaaaa taaagaagtg aaataaatgg 504

agaagatacc accttgatgg atggaagcct taaggtaaag atgctcattc tccccacact 564

gacctgtaca ttccccacag gcctaatcaa aacccccaca ggcttctgtg tagaaattga 624

catgctgatc ctgaaattta tatgaaaatg caaagagtct ggaataacca aaataatttt 684

gtaaaagaac aaagaagact tctactacct ggttataaga cttctctgaa gcacagaagt 744

caaggcagtg tggtggtggc ataagtaatg taaatcatcc aggtgtggtg gctcaagcca 804

gtcatccagc actttgggag gctgaggcag gagaatcgct agagcccagg agttggaaac 864

cagcctgggc agcatagc 882

<210> 13
<211> 442
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (58)..(336)

<400> 13
gctccggaat tcccgggaac atttacactg acagcaagta cgccttagct actgtgc 57
atg tac ata aag cca tct acc agg aaa tca ggc tac tca cct cag cag 105
Met Tyr Ile Lys Pro Ser Thr Arg Lys Ser Gly Tyr Ser Pro Gln Gln
1 5 10 15

gta gct gtg atc cac tgc aaa gga cat caa aaa gaa aac acg gcc gtt 153
Val Ala Val Ile His Cys Lys Gly His Gln Lys Glu Asn Thr Ala Val
20 25 30

gcc cat agt aac cag aaa gct gat tca gca gct cag gtc act gcc aga 201
Ala His Ser Asn Gln Lys Ala Asp Ser Ala Ala Gln Val Thr Ala Arg
35 40 45

ctt tca gtc acg cct cca aac ttg ctg ccc aca gtc tcc ttt cca cag 249
Leu Ser Val Thr Pro Pro Asn Leu Leu Pro Thr Val Ser Phe Pro Gln
50 55 60

cca gat ctg cct gac aat ccc gta tac tca aca aca aca gaa aaa ctg 297
Pro Asp Leu Pro Asp Asn Pro Val Tyr Ser Thr Thr Thr Glu Lys Leu
65 70 75 80

gct tca gat ctc aga gcc aat aaa aat cag gaa agt tag tagattcttc 346
Ala Ser Asp Leu Arg Ala Asn Lys Asn Gln Glu Ser *
85 90

ctgactctgg aatcttcata ccctgaactt aaaccagtta cctacagtct accacccatt 406

taagaagagc aaagttacct cagctcctcc ggaggg 442

<210> 14
<211> 2058
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (173)..(1909)

<400> 14
taagcttgcg gccgcccggc ggtgtagatg gtacgcttgc ttgtacgcag caggggaggc 60

tctgacttgc actggctggt ggtccgcttc cggggagagc gcccagtgcc cacaggtggc 120

```
aggagcctgt ccagggactc tgcattgctg ctgcagagag ggatgtactc gg    atg      175
                                                             Met
                                                              1

tgg ccc aga gct gcc cac ctc cca ggt acc cgc tgg cat ggc cgg ccc      223
Trp Pro Arg Ala Ala His Leu Pro Gly Thr Arg Trp His Gly Arg Pro
             5                  10                  15

aca ggg cat ccc ccc cca ccc cca atg agc ctg ccc gcc tca gcc atg      271
Thr Gly His Pro Pro Pro Pro Pro Met Ser Leu Pro Ala Ser Ala Met
         20                  25                  30

cct gtt gag ggg gtg ggg ggg gat gcc ctg tgg gcc ggc cat gcc agc      319
Pro Val Glu Gly Val Gly Gly Asp Ala Leu Trp Ala Gly His Ala Ser
     35                  40                  45

ggg tac ctg gga ggt ggc cag ctc tgg gcc aca tcc gag tac atc cct      367
Gly Tyr Leu Gly Gly Gly Gln Leu Trp Ala Thr Ser Glu Tyr Ile Pro
 50                  55                  60                  65

ctc tgc agc agc aat gca gag tcc ctg gac agg ctc ctg cca cct gtg      415
Leu Cys Ser Ser Asn Ala Glu Ser Leu Asp Arg Leu Leu Pro Pro Val
                 70                  75                  80

ggc act ggg cgc tct ccc cgg aag cgg acc acc agc cag tgc aag tca      463
Gly Thr Gly Arg Ser Pro Arg Lys Arg Thr Thr Ser Gln Cys Lys Ser
             85                  90                  95

gag cct ccc ctg ctg cgt aca agc aag cgt acc atc tac acc gcc ggg      511
Glu Pro Pro Leu Leu Arg Thr Ser Lys Arg Thr Ile Tyr Thr Ala Gly
        100                 105                 110

cgg ccg ccc tgg tac aat gaa cac ggc acg caa tcc aaa gag gcc ttc      559
Arg Pro Pro Trp Tyr Asn Glu His Gly Thr Gln Ser Lys Glu Ala Phe
    115                 120                 125

gcc atc ggc ttg gga ggc ggc agt gcc tct ggg aag acc act gtg gcc      607
Ala Ile Gly Leu Gly Gly Gly Ser Ala Ser Gly Lys Thr Thr Val Ala
130                 135                 140                 145

aga atg atc atc gag gcc ctg gat gtg ccc tgg gtg gtc ttg ctg tcc      655
Arg Met Ile Ile Glu Ala Leu Asp Val Pro Trp Val Val Leu Leu Ser
                150                 155                 160

atg gac tcc ttc tac aag gtg ctg act gag cag cag cag gaa cag gcc      703
Met Asp Ser Phe Tyr Lys Val Leu Thr Glu Gln Gln Gln Glu Gln Ala
            165                 170                 175

gca cac aac aac ttc aac ttc gac cac cca gat gcc ttt gac ttc gac      751
Ala His Asn Asn Phe Asn Phe Asp His Pro Asp Ala Phe Asp Phe Asp
        180                 185                 190

ctc atc att tcc acc ctc aag aag ctg aag cag ggg aag agt gtc aag      799
Leu Ile Ile Ser Thr Leu Lys Lys Leu Lys Gln Gly Lys Ser Val Lys
    195                 200                 205
```

```
gtg ccc att tat gac ttc acc acg cac agc cgg aag aag gac tgg aaa      847
Val Pro Ile Tyr Asp Phe Thr Thr His Ser Arg Lys Lys Asp Trp Lys
210                 215                 220                 225

aca ctg tat ggt gca aac gtc atc atc ttt gag ggc atc atg gcc ttt      895
Thr Leu Tyr Gly Ala Asn Val Ile Ile Phe Glu Gly Ile Met Ala Phe
                230                 235                 240

gct gac aag aca ctg ttg gag ctc ctg gac atg aag atc ttt gtg gac      943
Ala Asp Lys Thr Leu Leu Glu Leu Leu Asp Met Lys Ile Phe Val Asp
            245                 250                 255

aca gac tcc gac atc cgc ctg gta cgg cgg ctg cgc cgg gac atc agt      991
Thr Asp Ser Asp Ile Arg Leu Val Arg Arg Leu Arg Arg Asp Ile Ser
        260                 265                 270

gag cgc ggc cgg gac atc gag ggt gtc atc aag cag tac aac aag ttt     1039
Glu Arg Gly Arg Asp Ile Glu Gly Val Ile Lys Gln Tyr Asn Lys Phe
    275                 280                 285

gtc aag ccc tcc ttc gac cag tac atc cag ccc acc atg cgc ctg gca     1087
Val Lys Pro Ser Phe Asp Gln Tyr Ile Gln Pro Thr Met Arg Leu Ala
290                 295                 300                 305

gac atc gtg gtc ccc aga ggg agc ggc aac acg gtg gcc atc gac ctg     1135
Asp Ile Val Val Pro Arg Gly Ser Gly Asn Thr Val Ala Ile Asp Leu
                310                 315                 320

att gtg cag cac gtg cac agc cag ctg gag gag cgt gaa ctc agc gtc     1183
Ile Val Gln His Val His Ser Gln Leu Glu Glu Arg Glu Leu Ser Val
            325                 330                 335

agg gct gcg ctg gcc tcg gca cac cag tgc cac ccg ctg ccc cgg acg     1231
Arg Ala Ala Leu Ala Ser Ala His Gln Cys His Pro Leu Pro Arg Thr
        340                 345                 350

ctg agc gtc ctg aag agc acg ccg cag gta cgg ggc atg cac acc atc     1279
Leu Ser Val Leu Lys Ser Thr Pro Gln Val Arg Gly Met His Thr Ile
    355                 360                 365

atc agg gac aag gag acc agt cgc gac gag ttc atc ttc tac tcc aag     1327
Ile Arg Asp Lys Glu Thr Ser Arg Asp Glu Phe Ile Phe Tyr Ser Lys
370                 375                 380                 385

aga ctg atg cgg ctg ctc atc gag cac gcg ctc tcc ttc ctg ccc ttt     1375
Arg Leu Met Arg Leu Leu Ile Glu His Ala Leu Ser Phe Leu Pro Phe
                390                 395                 400

cag gac tgc gtc gta cag acc ccg cag ggg cag gac tat gcg ggc aag     1423
Gln Asp Cys Val Val Gln Thr Pro Gln Gly Gln Asp Tyr Ala Gly Lys
            405                 410                 415

tgc tat gcg ggg aag cag atc acc ggt gtg tcc att ctg cgc gcc ggt     1471
Cys Tyr Ala Gly Lys Gln Ile Thr Gly Val Ser Ile Leu Arg Ala Gly
        420                 425                 430

gaa acc atg gag ccc gcg ctg cgc gct gtg tgc aaa gac gtg cgc atc     1519
```

```
Glu Thr Met Glu Pro Ala Leu Arg Ala Val Cys Lys Asp Val Arg Ile
    435                 440                 445

ggc acc atc ctc atc cag acc aac cag ctt acc ggg gag ccc gag ctc     1567
Gly Thr Ile Leu Ile Gln Thr Asn Gln Leu Thr Gly Glu Pro Glu Leu
450                 455                 460                 465

cac tac ctg agg ctg ccc aag gac atc agc gat gac cac gtg atc ctc     1615
His Tyr Leu Arg Leu Pro Lys Asp Ile Ser Asp Asp His Val Ile Leu
                470                 475                 480

atg gac tgc acc gtg tcc acg ggc gcg gcg gcc atg atg gca gtg cgc     1663
Met Asp Cys Thr Val Ser Thr Gly Ala Ala Ala Met Met Ala Val Arg
            485                 490                 495

gtg ctc ctg gac cac gac gtg cct gag gac aag atc ttt ttg ctg tcg     1711
Val Leu Leu Asp His Asp Val Pro Glu Asp Lys Ile Phe Leu Leu Ser
        500                 505                 510

ctg ctc atg gca gag atg ggc gtg cac tca gtg gcc tat gca ttt ccg     1759
Leu Leu Met Ala Glu Met Gly Val His Ser Val Ala Tyr Ala Phe Pro
    515                 520                 525

cga gtg aga atc atc acc acg gcg gtg gac aag cgg gtc aat gac ctt     1807
Arg Val Arg Ile Ile Thr Thr Ala Val Asp Lys Arg Val Asn Asp Leu
530                 535                 540                 545

ttc cgc atc atc cca ggc att ggg aac ttt ggc gac cgc tac ttt ggg     1855
Phe Arg Ile Ile Pro Gly Ile Gly Asn Phe Gly Asp Arg Tyr Phe Gly
                550                 555                 560

aca gac gcg gtc ccc gat ggc agt gac gag gag gaa gtg gcc tac acg     1903
Thr Asp Ala Val Pro Asp Gly Ser Asp Glu Glu Glu Val Ala Tyr Thr
            565                 570                 575

ggt tag ctgcccagtg agccatcccg tccccaccac cctcctcctg cctcctgacc     1959
Gly  *


caggactgct gaatacaaag atgttaattt ttaaaatgtt actagtataa tttattctat   2019

gcattttata aaataaataa agctttagaa aaaaaaaaa                          2058
```

<210> 15
<211> 1705
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (545)..(1672)

<400> 15
aaggatcctt aattttatta atccccccc ccccgaaggc ccccagccgc ctgtgccccg   60

```
cggtgcgcgc cccgctccgg gacctgccgc caccgccgcc ccgccctcgg cggcacccac    120

acccaggcgc gcccgcgcgc gcgcccggcc ccgtccctgc ctggaagcac agctgaagat    180

ggcgagcccg gcgcctccgg agcacgccga ggagggatgc ccggctcctg ccgccgagga    240

gcaggcgccg ccgtcgccgc caccgcccca ggcatccccc gcagagcggc agcagcagga    300

ggaggaagcg caggaagctg gggcggcgga gggcgcgggg ttgcaggtgg aggaggccgc    360

gggccgggcg gcggccgcgg taacctggct gctcggggag ccggtgctgt ggctgggctg    420

ccgcgccgac gagctgctga gctggaagag gccgctgcgg agcctgctcg gcttcgtcgc    480

tgccaacctg ctgttctggt tccttgcatt gactccatgg agagtatatc acctgatttc    540

cgtc atg ata ctt ggg cgt gtt att atg caa ata ata aag gat atg gtt      589
     Met Ile Leu Gly Arg Val Ile Met Gln Ile Ile Lys Asp Met Val
       1               5                  10                  15

ttg tct aga aca aga ggt gca cag ttg tgg aga agc ctc agt gaa agc      637
Leu Ser Arg Thr Arg Gly Ala Gln Leu Trp Arg Ser Leu Ser Glu Ser
                20                  25                  30

tgg gaa gtt atc aat tcc aaa cca gat gaa aga ccc agg ctc agc cac      685
Trp Glu Val Ile Asn Ser Lys Pro Asp Glu Arg Pro Arg Leu Ser His
            35                  40                  45

tgt att gca gaa tca tgg atg aat ttc agc ata ttt ctt caa gaa atg      733
Cys Ile Ala Glu Ser Trp Met Asn Phe Ser Ile Phe Leu Gln Glu Met
        50                  55                  60

tct ctt ttt aaa cag cag agc cct ggc aag ttt tgt ctc ctg gtc tgt      781
Ser Leu Phe Lys Gln Gln Ser Pro Gly Lys Phe Cys Leu Leu Val Cys
    65                  70                  75

agt gtg tgc aca ttt ttt acg atc ttg gga agt tac att cct ggg gtt      829
Ser Val Cys Thr Phe Phe Thr Ile Leu Gly Ser Tyr Ile Pro Gly Val
80                  85                  90                  95

ata ctc agc tat cta ctg tta ctg tgt gca ttt ttg tgt cca ttg ttt      877
Ile Leu Ser Tyr Leu Leu Leu Leu Cys Ala Phe Leu Cys Pro Leu Phe
                100                 105                 110

aaa tgt aat gat att gga caa aaa att tac agc aaa att aag tca gtt      925
Lys Cys Asn Asp Ile Gly Gln Lys Ile Tyr Ser Lys Ile Lys Ser Val
            115                 120                 125

ctg ctg aaa ctg gat ttt gga att gga gaa tat att aat cag aag aaa      973
Leu Leu Lys Leu Asp Phe Gly Ile Gly Glu Tyr Ile Asn Gln Lys Lys
        130                 135                 140

cgt gag aga tct gaa gca gat aaa gaa aaa agt cac aaa gat gac agt     1021
Arg Glu Arg Ser Glu Ala Asp Lys Glu Lys Ser His Lys Asp Asp Ser
    145                 150                 155
```

```
gaa tta gac ttt tca gct ctt tgt cct aag att agc ctc acg gtt gct     1069
Glu Leu Asp Phe Ser Ala Leu Cys Pro Lys Ile Ser Leu Thr Val Ala
160                 165                 170                 175

gcc aaa gag tta tct gtg tct gac aca gac gtc tca gag gta tcc tgg     1117
Ala Lys Glu Leu Ser Val Ser Asp Thr Asp Val Ser Glu Val Ser Trp
                180                 185                 190

act gat aat ggg acc ttc aac ctt tca gaa gga tac act cca cag aca     1165
Thr Asp Asn Gly Thr Phe Asn Leu Ser Glu Gly Tyr Thr Pro Gln Thr
            195                 200                 205

gac act tct gat gat ctt gac cga ccc agt gag gaa gtt ttc tct aga     1213
Asp Thr Ser Asp Asp Leu Asp Arg Pro Ser Glu Glu Val Phe Ser Arg
        210                 215                 220

gat ctt tca gat ttt cca tct cta gaa aat ggc atg gga aca aat gat     1261
Asp Leu Ser Asp Phe Pro Ser Leu Glu Asn Gly Met Gly Thr Asn Asp
    225                 230                 235

gaa gat gaa tta agc ctt ggt ttg ccc act gag ctc aag aga aag aag     1309
Glu Asp Glu Leu Ser Leu Gly Leu Pro Thr Glu Leu Lys Arg Lys Lys
240                 245                 250                 255

gaa cag ttg gac agt ggt cac aga cca agc aaa gag acg caa tca gca     1357
Glu Gln Leu Asp Ser Gly His Arg Pro Ser Lys Glu Thr Gln Ser Ala
                260                 265                 270

gct ggt ctc acc ctt cct ctg aac agt gac caa acc ttt cac ctg atg     1405
Ala Gly Leu Thr Leu Pro Leu Asn Ser Asp Gln Thr Phe His Leu Met
            275                 280                 285

agc aac ctg gct ggg gat gtt atc aca gct gca gtg act gca gct atc     1453
Ser Asn Leu Ala Gly Asp Val Ile Thr Ala Ala Val Thr Ala Ala Ile
        290                 295                 300

aaa gac cag tta gag ggt gtg cag caa gca ctt tct cag gct gcc ccc     1501
Lys Asp Gln Leu Glu Gly Val Gln Gln Ala Leu Ser Gln Ala Ala Pro
    305                 310                 315

atc cca gaa gag gac aca gac act gaa gaa ggt gat gac ttt gaa cta     1549
Ile Pro Glu Glu Asp Thr Asp Thr Glu Glu Gly Asp Asp Phe Glu Leu
320                 325                 330                 335

ctt gac cag tca gag ctg gat caa att gag agt gaa ttg gga ctt aca     1597
Leu Asp Gln Ser Glu Leu Asp Gln Ile Glu Ser Glu Leu Gly Leu Thr
                340                 345                 350

caa gac cag gaa gca gaa gca cag caa aat aag aag tct tca ggt ttc     1645
Gln Asp Gln Glu Ala Glu Ala Gln Gln Asn Lys Lys Ser Ser Gly Phe
            355                 360                 365

ctt tca aat ctg ctg gga ggc cat taa tctag gaatcagctt gcaacagagc     1697
Leu Ser Asn Leu Leu Gly Gly His  *
        370                 375

acaaaaaa                                                            1705
```

<210> 16
<211> 1914
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (154)..(1593)

<400> 16
gggagggtgt atggatgata acgctagtac catccggcct atttaggtga cactatagaa 60

caagtttgta caaaaaagca ggctggtacc ggtccggaat tcccgggata tcgtcgaccc 120

acgcgtccgg tgcaccgcgt tctcgcacgc gtc atg gcg gtc ctc gga gta cag 174
Met Ala Val Leu Gly Val Gln
1 5

ctg gtg gtg acc ctg ctc act gcc acc ctc atg cac agg ctg gcg cca 222
Leu Val Val Thr Leu Leu Thr Ala Thr Leu Met His Arg Leu Ala Pro
10 15 20

cac tgc tcc ttc gcg cgc tgg ctg ctc tgt aac ggc agt ttg ttc cga 270
His Cys Ser Phe Ala Arg Trp Leu Leu Cys Asn Gly Ser Leu Phe Arg
25 30 35

tac aag cac ccg tct gag gag gag ctt cgg gcc ctg gcg ggg aag ccg 318
Tyr Lys His Pro Ser Glu Glu Glu Leu Arg Ala Leu Ala Gly Lys Pro
40 45 50 55

agg ccc aga ggc agg aaa gag cgg tgg gcc aat ggc ctt agt gag gag 366
Arg Pro Arg Gly Arg Lys Glu Arg Trp Ala Asn Gly Leu Ser Glu Glu
60 65 70

aag cca ctg tct gtg ccc cga gat gcc ccg ttc cag ctg gag acc tgc 414
Lys Pro Leu Ser Val Pro Arg Asp Ala Pro Phe Gln Leu Glu Thr Cys
75 80 85

ccc ctc acg acc gtg gat gcc ctg gtc ctg cgc ttc ttc ctg gag tac 462
Pro Leu Thr Thr Val Asp Ala Leu Val Leu Arg Phe Phe Leu Glu Tyr
90 95 100

cag tgg ttt gtg gac ttt gct gtg tac tcg ggc ggc gtg tac ctc ttc 510
Gln Trp Phe Val Asp Phe Ala Val Tyr Ser Gly Gly Val Tyr Leu Phe
105 110 115

aca gag gcc tac tac tac atg ctg gga cca gcc aag gag act aac att 558
Thr Glu Ala Tyr Tyr Tyr Met Leu Gly Pro Ala Lys Glu Thr Asn Ile
120 125 130 135

gct gtg ttc tgg tgc ctg ctc acg gtg acc ttc tcc atc aag atg ttc 606
Ala Val Phe Trp Cys Leu Leu Thr Val Thr Phe Ser Ile Lys Met Phe
140 145 150

```
ctg aca gtg aca cgg ctg tac ttc agc gcc gag gag ggg ggt gag cgc      654
Leu Thr Val Thr Arg Leu Tyr Phe Ser Ala Glu Glu Gly Gly Glu Arg
            155                 160                 165

tct gtc tgc ctc acc ttt gcc ttc ctc ttc ctg ctg ctg gcc atg ctg      702
Ser Val Cys Leu Thr Phe Ala Phe Leu Phe Leu Leu Leu Ala Met Leu
        170                 175                 180

gtg caa gtg gtg cgg gag gag acc ctc gag ctg ggc ctg gag cct ggt      750
Val Gln Val Val Arg Glu Glu Thr Leu Glu Leu Gly Leu Glu Pro Gly
    185                 190                 195

ctg gcc agc atg acc cag aac tta gag cca ctt ctg aag aag cag ggc      798
Leu Ala Ser Met Thr Gln Asn Leu Glu Pro Leu Leu Lys Lys Gln Gly
200                 205                 210                 215

tgg gac tgg gcg ctt cct gtg gcc aag ctg gct atc cgc gtg gga ctg      846
Trp Asp Trp Ala Leu Pro Val Ala Lys Leu Ala Ile Arg Val Gly Leu
                220                 225                 230

gca gtg gtg ggc tct gtg ctg ggt gcc ttc ctc acc ttc cca ggc ctg      894
Ala Val Val Gly Ser Val Leu Gly Ala Phe Leu Thr Phe Pro Gly Leu
            235                 240                 245

cgg ctg gcc cag acc cac cgg gac gca ctg acc atg tcg gag gac aga      942
Arg Leu Ala Gln Thr His Arg Asp Ala Leu Thr Met Ser Glu Asp Arg
        250                 255                 260

ccc atg ctg cag ttc ctc ctg cac acc agc ttc ctg tct ccc ctg ttc      990
Pro Met Leu Gln Phe Leu Leu His Thr Ser Phe Leu Ser Pro Leu Phe
    265                 270                 275

atc ctg tgg ctc tgg aca aag ccc att gca cgg gac ttc ctg cac cag     1038
Ile Leu Trp Leu Trp Thr Lys Pro Ile Ala Arg Asp Phe Leu His Gln
280                 285                 290                 295

ccg ccg ttt ggg gag acg cgt ttc tcc ctg ctg tcc gat tct gcc ttc     1086
Pro Pro Phe Gly Glu Thr Arg Phe Ser Leu Leu Ser Asp Ser Ala Phe
                300                 305                 310

gac tct ggg cgc ctc tgg ttg ctg gtg gtg ctg tgc ctg ctg cgg ctg     1134
Asp Ser Gly Arg Leu Trp Leu Leu Val Val Leu Cys Leu Leu Arg Leu
            315                 320                 325

gcg gtg acc cgg ccc cac ctg cag gcc tac ctg tgc ctg gcc aag gcc     1182
Ala Val Thr Arg Pro His Leu Gln Ala Tyr Leu Cys Leu Ala Lys Ala
        330                 335                 340

cgg gtg gag cag ctg cga agg gag gct ggc cgc atc gaa gcc cgt gaa     1230
Arg Val Glu Gln Leu Arg Arg Glu Ala Gly Arg Ile Glu Ala Arg Glu
    345                 350                 355

atc cag cag agg gtg gtc cga gtc tac tgc tat gtg acc gtg gtg agc     1278
Ile Gln Gln Arg Val Val Arg Val Tyr Cys Tyr Val Thr Val Val Ser
360                 365                 370                 375
```

```
ttg cag tac ctg acg ccg ctc atc ctc acc ctc aac tgc aca ctt ctg     1326
Leu Gln Tyr Leu Thr Pro Leu Ile Leu Thr Leu Asn Cys Thr Leu Leu
                380                 385                 390

ctc aag acg ctg gga ggc tat tcc tgg ggc ctg ggc cca gct cct cta     1374
Leu Lys Thr Leu Gly Gly Tyr Ser Trp Gly Leu Gly Pro Ala Pro Leu
            395                 400                 405

cta tcc ccc gac cca tcc tca gcc agc gct gcc ccc atc ggc tct ggg     1422
Leu Ser Pro Asp Pro Ser Ser Ala Ser Ala Ala Pro Ile Gly Ser Gly
        410                 415                 420

gag gac gaa gtc cag cag act gca gcg cgg att gcc ggg gct ctg ggt     1470
Glu Asp Glu Val Gln Gln Thr Ala Ala Arg Ile Ala Gly Ala Leu Gly
    425                 430                 435

ggc ctg ctt act ccc ctc ttc ctc cgt ggc gtc ctg gcc tac ctc atc     1518
Gly Leu Leu Thr Pro Leu Phe Leu Arg Gly Val Leu Ala Tyr Leu Ile
440                 445                 450                 455

tgg tgg acg gct gcc tgc cag ctg ctc gcc agc ctt ttc ggc ctc tac     1566
Trp Trp Thr Ala Ala Cys Gln Leu Leu Ala Ser Leu Phe Gly Leu Tyr
                460                 465                 470

ttc cac cag cac ttg gca ggc tcc tag ctgcc tgcagaccct cctggggccc     1618
Phe His Gln His Leu Ala Gly Ser  *
            475                 480

tgaggtctgt tcctggggca gcgggacact agcctgcccc ctctgtttgc gcccccgtgt   1678

ccccagctgc aaggtggggc cggactcccc ggcgttccct tcaccacagt gcctgacccg   1738

cggcccccct tggacgccga gtttctgcct cagaactgtc tctcctgggc ccagcagcat   1798

gagggtcccg aggccattgt ctccgaagcg tatgtgccag gtttgagtgg cgagggtgat   1858

gctggctgct cttctgaaca aataaaggag catgccgatt tttacaaaaa aaaaaa       1914
```

<210> 17
<211> 859
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (61)..(795)

<400> 17

```
atttggccct cgaggccaag aattcggcac gaggaaggag gtcccccacg gcccttcagg     60

atg aaa gct gcg gtg ctg acc ttg gcc gtg ctc ttc ctg acg ggg agc      108
Met Lys Ala Ala Val Leu Thr Leu Ala Val Leu Phe Leu Thr Gly Ser
  1               5                  10                  15
```

```
cag gct cgg cat ttc tgg cag caa gat gaa ccc ccc cag agc ccc tgg      156
Gln Ala Arg His Phe Trp Gln Gln Asp Glu Pro Pro Gln Ser Pro Trp
            20                  25                  30

gat cga gtg aag gac ctg gcc act gtg tac gtg gat gtg ctc aaa gac      204
Asp Arg Val Lys Asp Leu Ala Thr Val Tyr Val Asp Val Leu Lys Asp
        35                  40                  45

agc ggc aaa gac agc gtg acc tcc acc ttc agc aag ctg cgc gaa cag      252
Ser Gly Lys Asp Ser Val Thr Ser Thr Phe Ser Lys Leu Arg Glu Gln
    50                  55                  60

ctc ggc cct gtg acc cag gag ttc tgg gat aac ctg gaa aag gag aca      300
Leu Gly Pro Val Thr Gln Glu Phe Trp Asp Asn Leu Glu Lys Glu Thr
65                  70                  75                  80

gag ggc ctg agg cag gag atg agc aag gat ctg gag gag gtg aag gcc      348
Glu Gly Leu Arg Gln Glu Met Ser Lys Asp Leu Glu Glu Val Lys Ala
                85                  90                  95

aag gtg cag ccc tac ctg gac gac ttc cag aag aag tgg cag gag gag      396
Lys Val Gln Pro Tyr Leu Asp Asp Phe Gln Lys Lys Trp Gln Glu Glu
            100                 105                 110

atg gag ctc tac cgc cag aag gtg gag ccg ctg cgc gca gag ctc caa      444
Met Glu Leu Tyr Arg Gln Lys Val Glu Pro Leu Arg Ala Glu Leu Gln
        115                 120                 125

gag ggc gcg cgc cag aag ctg cac gag ctg caa gag aag ctg agc cca      492
Glu Gly Ala Arg Gln Lys Leu His Glu Leu Gln Glu Lys Leu Ser Pro
    130                 135                 140

ctg ggc gag gag atg cgc gac cgc gcg cgc gcc cat gtg gac gcg ctg      540
Leu Gly Glu Glu Met Arg Asp Arg Ala Arg Ala His Val Asp Ala Leu
145                 150                 155                 160

cgc acg cat ctg gcc ccc tac agc gac gag ctg cgc cag cgc ttg gcc      588
Arg Thr His Leu Ala Pro Tyr Ser Asp Glu Leu Arg Gln Arg Leu Ala
                165                 170                 175

gcg cgc ctt gag gct ctc aag gag aac ggc ggc gcc aga ctg gcc gag      636
Ala Arg Leu Glu Ala Leu Lys Glu Asn Gly Gly Ala Arg Leu Ala Glu
            180                 185                 190

tac cac gcc aag gcc acc gag cat ctg agc acg ctc agc gag aag gcc      684
Tyr His Ala Lys Ala Thr Glu His Leu Ser Thr Leu Ser Glu Lys Ala
        195                 200                 205

aag ccc gcg ctc gag gac ctc cgc caa ggc ctg ctg ccc gtg ctg gag      732
Lys Pro Ala Leu Glu Asp Leu Arg Gln Gly Leu Leu Pro Val Leu Glu
    210                 215                 220

agc ttc aag gtc agc ttc ctg agc gct ctc gag gag tac act aag aag      780
Ser Phe Lys Val Ser Phe Leu Ser Ala Leu Glu Glu Tyr Thr Lys Lys
225                 230                 235                 240

ctc aac acc cag tga ggcgcccgcc gccgcccccc ttcccggtgc tcagaataaa      835
```

```
Leu Asn Thr Gln  *
                245

cgtttccaaa gtgggaaaaa aaaa                                              859



     <210> 18
     <211> 2454
     <212> DNA
     <213> Homo sapiens

     <220>
     <221> CDS
     <222> (154)..(1710)

     <400> 18
gtgaaaggag ggaacgcagg tgagaaagcg agacaggcag gtagggaaat cgtgaggtga      60

gcgtgatcct agctccttgt ggcagagcct agagagaagg cgaggacgct gaagaaccag     120

gcggacagct ggcagagaga gaagttggct agc atg gaa tca cca gag gag cct      174
                                    Met Glu Ser Pro Glu Glu Pro
                                      1               5

gga gca tcc atg gat gag aac tac ttt gtg aac tac act ttc aaa gat       222
Gly Ala Ser Met Asp Glu Asn Tyr Phe Val Asn Tyr Thr Phe Lys Asp
         10                  15                  20

cgg tca cat tca ggc cgt gtg gct caa ggc atc atg aaa ctg tgt cta       270
Arg Ser His Ser Gly Arg Val Ala Gln Gly Ile Met Lys Leu Cys Leu
     25                  30                  35

gag gag gag ctc ttt gct gat gtc acc att tcg gtg gaa ggc cgg gag       318
Glu Glu Glu Leu Phe Ala Asp Val Thr Ile Ser Val Glu Gly Arg Glu
 40                  45                  50                  55

ttt cag ctc cat cgg ctg gtc ctc tca gct cag agc tgc ttc ttc cga       366
Phe Gln Leu His Arg Leu Val Leu Ser Ala Gln Ser Cys Phe Phe Arg
                 60                  65                  70

tcc atg ttc act tcc aac ctg aag gag gcc cac aac cgg gtg att gtg       414
Ser Met Phe Thr Ser Asn Leu Lys Glu Ala His Asn Arg Val Ile Val
             75                  80                  85

ctg cag gat gtc agc gag tct gtt ttc cag ctc ctg gtt gat tat atc       462
Leu Gln Asp Val Ser Glu Ser Val Phe Gln Leu Leu Val Asp Tyr Ile
         90                  95                 100

tac cat ggg act gtg aaa ctt cga gct gag gag ttg cag gaa att tat       510
Tyr His Gly Thr Val Lys Leu Arg Ala Glu Glu Leu Gln Glu Ile Tyr
    105                 110                 115

gag gtg tca gac atg tat cag ctg aca tct ctc ttt gag gaa tgc tct       558
Glu Val Ser Asp Met Tyr Gln Leu Thr Ser Leu Phe Glu Glu Cys Ser
120                 125                 130                 135
```

```
cgg ttt ttg gcc cgc aca gtg caa gtg gga aac tgc ctt cag gtg atg      606
Arg Phe Leu Ala Arg Thr Val Gln Val Gly Asn Cys Leu Gln Val Met
                140                 145                 150

tgg ctg gca gat cgg cac agt gat cct gag ctc tat acg gct gcc aag      654
Trp Leu Ala Asp Arg His Ser Asp Pro Glu Leu Tyr Thr Ala Ala Lys
            155                 160                 165

cac tgt gcc aag acc cac ctg gcc cag ctg cag aat aca gag gaa ttt      702
His Cys Ala Lys Thr His Leu Ala Gln Leu Gln Asn Thr Glu Glu Phe
        170                 175                 180

ctc cac ttg ccc cac cgc tta ctc aca gat atc atc tcg gat gga gtt      750
Leu His Leu Pro His Arg Leu Leu Thr Asp Ile Ile Ser Asp Gly Val
    185                 190                 195

ccg tgt tct cag aac cca aca gag gca ata gaa gcc tgg atc aac ttt      798
Pro Cys Ser Gln Asn Pro Thr Glu Ala Ile Glu Ala Trp Ile Asn Phe
200                 205                 210                 215

aat aaa gag gaa aga gag gct ttt gca gag tca ctc agg aca agc ttg      846
Asn Lys Glu Glu Arg Glu Ala Phe Ala Glu Ser Leu Arg Thr Ser Leu
                220                 225                 230

aag gaa att ggg gag aat gtg cac att tac ctg att ggg aaa gag tca      894
Lys Glu Ile Gly Glu Asn Val His Ile Tyr Leu Ile Gly Lys Glu Ser
            235                 240                 245

tct cgt acc cac tcg ttg gct gtg tcc ttg cac tgt gca gaa gat gac      942
Ser Arg Thr His Ser Leu Ala Val Ser Leu His Cys Ala Glu Asp Asp
        250                 255                 260

tcc atc agt gta agt ggc caa aac agt ttg tgc cac cag atc act gcg      990
Ser Ile Ser Val Ser Gly Gln Asn Ser Leu Cys His Gln Ile Thr Ala
    265                 270                 275

gcc tgc aag cat ggt gga gac ttg tat gtg gtg gga ggg tcc atc cca     1038
Ala Cys Lys His Gly Gly Asp Leu Tyr Val Val Gly Gly Ser Ile Pro
280                 285                 290                 295

cgg cgc atg tgg aag tgc aac aat gcc acc gtt gac tgg gag tgg tgt     1086
Arg Arg Met Trp Lys Cys Asn Asn Ala Thr Val Asp Trp Glu Trp Cys
                300                 305                 310

gct cct ttg cct cgg gac cgg ctc cag cac acc ctg gtg tct gtg ccc     1134
Ala Pro Leu Pro Arg Asp Arg Leu Gln His Thr Leu Val Ser Val Pro
            315                 320                 325

ggg aaa gat gcc ata tat tca ctg ggt ggc aag aca ctg caa gat acc     1182
Gly Lys Asp Ala Ile Tyr Ser Leu Gly Gly Lys Thr Leu Gln Asp Thr
        330                 335                 340

ctc tcc aac gca gtc att tat tat cgc gta ggt gat aat gtg tgg aca     1230
Leu Ser Asn Ala Val Ile Tyr Tyr Arg Val Gly Asp Asn Val Trp Thr
    345                 350                 355
```

```
gag aca act cag cta gag gtg gct gtg tca ggg gct gct ggt gcc aac     1278
Glu Thr Thr Gln Leu Glu Val Ala Val Ser Gly Ala Ala Gly Ala Asn
360                 365                 370                 375

ctc aac ggg atc atc tac tta cta ggg ggg gag gag aat gat ctg gac     1326
Leu Asn Gly Ile Ile Tyr Leu Leu Gly Gly Glu Glu Asn Asp Leu Asp
                380                 385                 390

ttc ttt acc aaa cct tcc cga ctc atc cag tgc ttt gac aca gag aca     1374
Phe Phe Thr Lys Pro Ser Arg Leu Ile Gln Cys Phe Asp Thr Glu Thr
            395                 400                 405

gac aaa tgc cat gtg aag ccc tat gtg ctg ccc ttt gca ggc cgc atg     1422
Asp Lys Cys His Val Lys Pro Tyr Val Leu Pro Phe Ala Gly Arg Met
        410                 415                 420

cac gca gct gtg cat aaa gat ctg gtg ttc atc gtg gct gaa ggg gac     1470
His Ala Ala Val His Lys Asp Leu Val Phe Ile Val Ala Glu Gly Asp
    425                 430                 435

tcc ctg gtg tgc tac aat ccc ttg cta gac agc ttc acc cgg ctt tgc     1518
Ser Leu Val Cys Tyr Asn Pro Leu Leu Asp Ser Phe Thr Arg Leu Cys
440                 445                 450                 455

ctt cct gag gcc tgg agc tct gcc cca tcc ctc tgg aag att gcc agc     1566
Leu Pro Glu Ala Trp Ser Ser Ala Pro Ser Leu Trp Lys Ile Ala Ser
                460                 465                 470

tgt aac ggg agc atc tat gtc ttc cgg gac cga tat aaa aag ggg gat     1614
Cys Asn Gly Ser Ile Tyr Val Phe Arg Asp Arg Tyr Lys Lys Gly Asp
            475                 480                 485

gcc aac acc tac aag ctt gac cct gcc act tca gcc gta act gtc aca     1662
Ala Asn Thr Tyr Lys Leu Asp Pro Ala Thr Ser Ala Val Thr Val Thr
        490                 495                 500

aga ggt att aag gtg ctg ctt acc aat ttg cag ttt gtg ttg gcc taa     1710
Arg Gly Ile Lys Val Leu Leu Thr Asn Leu Gln Phe Val Leu Ala  *
    505                 510                 515

ggctgtgggg aggggaggag aactgctcac tccttttccc tccccataca aactcaaagt   1770

cccctgggcc ccaattcaga gttatgtttt ttttggcaca tactagaaag gcagtgcctc   1830

agcccttccc tgaatccatg gaggtgttct gtttggggct ttttagactg ctgctgctca   1890

gctggttgct tgaactgaca gtaggccagc ctgttctctg ccattcccta gtcatcctgt   1950

gcctcaccac agcttgctta gagcaagcct tttctcagac cttaggcaca gcctctcctc   2010

tttacctgat caatgttaaa tgtaagcacc cctgatccca ggacataagg aaagatgccc   2070

aattgtactt ttgttctata gcctgtgaaa tggctagttg atcatttttc cacaaagaat   2130

taggtgttaa gagttttcct tcaggcttta cttaggagaa tggactaagc tgaaggtgta   2190

cttcaccagc aagagtcaac tctagaattc aggatgttcc ttctattgtt ttcttagcca   2250
```

```
tctgtcagga aatgtaactt tggttttatt tttggcttat tccaaggggt aagccagaaa   2310

atagaaatga ttatttctga ttaatagcag aaacttttтc aatctcaaat atataaggtg   2370

tctgctcttt taaaagctct aagctaagtc aagagctagg aactgttgat acaaataaaa   2430

gtttttgaag ggaaaaaaaa aaaa                                          2454
```

<210> 19
<211> 3586
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (106)..(1860)

<400> 19

```
gtaccggtcc ggaattcccg ggtcgacgat ttcgtggagc aggcagacgt cgaagccgtc     60

ctgcagccga ccctcctgca tctgggcacg ttcaggttgc gacac    atg aag gct     114
                                                     Met Lys Ala
                                                       1

ttg atg aac gag aag gcc cag gcc gcc ctg gtg gag ttt gtg gag gac      162
Leu Met Asn Glu Lys Ala Gln Ala Ala Leu Val Glu Phe Val Glu Asp
  5                  10                  15

gtc aat cac gct gcc att ccc agg gag atc cca cgc aag gat ggg gtc      210
Val Asn His Ala Ala Ile Pro Arg Glu Ile Pro Arg Lys Asp Gly Val
 20                  25                  30                  35

tgg agg gtt ctg tgg aag gac cgt gcg cag gac acg agg gtc ctg agg      258
Trp Arg Val Leu Trp Lys Asp Arg Ala Gln Asp Thr Arg Val Leu Arg
                 40                  45                  50

cag atg aca cgc ctg ctg ctg gat gac ggg ccc acg cag gcc gcg gag      306
Gln Met Thr Arg Leu Leu Leu Asp Asp Gly Pro Thr Gln Ala Ala Glu
             55                  60                  65

gct ggg acc ccc ggt gag gca ccc acc cct ccc gct tcg gag acg cag      354
Ala Gly Thr Pro Gly Glu Ala Pro Thr Pro Pro Ala Ser Glu Thr Gln
         70                  75                  80

gcc cag gat tct ggt gag gta aca ggg cat gct ggc tcg ctt ctt ggg      402
Ala Gln Asp Ser Gly Glu Val Thr Gly His Ala Gly Ser Leu Leu Gly
     85                  90                  95

gca ccc agg aac cca agg agg ggc cgt cgg ggt cgc aga aac aga acc      450
Ala Pro Arg Asn Pro Arg Arg Gly Arg Arg Gly Arg Arg Asn Arg Thr
100                 105                 110                 115

aga cgc aac agg ttg acc cag aag ggc aag aag aga agc cga gga gga      498
```

```
Arg Arg Asn Arg Leu Thr Gln Lys Gly Lys Lys Arg Ser Arg Gly Gly
                120                 125                 130

cgg ccg tct gct ccc gcg agg agt gag gcc gag gac tct tcc gac gag      546
Arg Pro Ser Ala Pro Ala Arg Ser Glu Ala Glu Asp Ser Ser Asp Glu
            135                 140                 145

agc ctg ggc atc gtg atc gag gag atc gac cag ggc gac ctg agc gga      594
Ser Leu Gly Ile Val Ile Glu Glu Ile Asp Gln Gly Asp Leu Ser Gly
        150                 155                 160

gaa gag gac cag agc gcg ctg tac gcc acg ctg cag gcc gct gcc agg      642
Glu Glu Asp Gln Ser Ala Leu Tyr Ala Thr Leu Gln Ala Ala Ala Arg
    165                 170                 175

gag ctg gtt agg cag tgg gcg ccc tgc aac tcc gag ggg gaa gaa gac      690
Glu Leu Val Arg Gln Trp Ala Pro Cys Asn Ser Glu Gly Glu Glu Asp
180                 185                 190                 195

ggt ccc cgc gag ttc ttg gct ctg gtc acc gtc acc gac aaa tcg aag      738
Gly Pro Arg Glu Phe Leu Ala Leu Val Thr Val Thr Asp Lys Ser Lys
                200                 205                 210

aaa gaa gag gca gag aag gag cca gct ggg gcc gaa tcc atc cgc ttg      786
Lys Glu Glu Ala Glu Lys Glu Pro Ala Gly Ala Glu Ser Ile Arg Leu
            215                 220                 225

aac acc aaa gaa gac aaa aat ggt gtc ccc gac tta gtg gcc ctg ctg      834
Asn Thr Lys Glu Asp Lys Asn Gly Val Pro Asp Leu Val Ala Leu Leu
        230                 235                 240

gct gtg aga gac acc ccg gac gag gag ccg gtg gac agc gac act tcg      882
Ala Val Arg Asp Thr Pro Asp Glu Glu Pro Val Asp Ser Asp Thr Ser
    245                 250                 255

gag agc gac tcg cag gaa agt ggg gac caa gaa aca gag gag ttg gat      930
Glu Ser Asp Ser Gln Glu Ser Gly Asp Gln Glu Thr Glu Glu Leu Asp
260                 265                 270                 275

aat cct gag ttc gtg gcc att gtg gcc tat acc gac ccg tcg gac ccc      978
Asn Pro Glu Phe Val Ala Ile Val Ala Tyr Thr Asp Pro Ser Asp Pro
                280                 285                 290

tgg gcc cgg gag gag atg ttg aaa atc gct tct gtt atc gag tcg ctg     1026
Trp Ala Arg Glu Glu Met Leu Lys Ile Ala Ser Val Ile Glu Ser Leu
            295                 300                 305

ggc tgg agc gac gag aaa gac aag cga gac ccc ctc cga cag gtc ttg     1074
Gly Trp Ser Asp Glu Lys Asp Lys Arg Asp Pro Leu Arg Gln Val Leu
        310                 315                 320

tcc gtc atg tcc aag gac act aac ggg acc cgc gtg aag gtg gaa gag     1122
Ser Val Met Ser Lys Asp Thr Asn Gly Thr Arg Val Lys Val Glu Glu
    325                 330                 335

gcg ggc cgc gag gtg gac gcc gtg gtc ctg cgc aag gcc ggg gat gac     1170
Ala Gly Arg Glu Val Asp Ala Val Val Leu Arg Lys Ala Gly Asp Asp
```

```
340                 345                 350                 355

ggg gac ctc cgg gag tgc att tcc acc ttg gcg cag ccg gat ctc cct     1218
Gly Asp Leu Arg Glu Cys Ile Ser Thr Leu Ala Gln Pro Asp Leu Pro
                360                 365                 370

ccc cag gcg aag aag gct ggg cgt ggc ctc ttc ggg ggc tgg agc gag     1266
Pro Gln Ala Lys Lys Ala Gly Arg Gly Leu Phe Gly Gly Trp Ser Glu
            375                 380                 385

cac cgt gag gac gaa ggg ggt ctt ctg gag ctg gtg gcg ctc ctg gct     1314
His Arg Glu Asp Glu Gly Gly Leu Leu Glu Leu Val Ala Leu Leu Ala
        390                 395                 400

gcc cag gac atg gcg gag gtg atg aag gag gaa aaa gaa aac gcc tgg     1362
Ala Gln Asp Met Ala Glu Val Met Lys Glu Glu Lys Glu Asn Ala Trp
    405                 410                 415

gaa ggc ggg aag tac aaa tac ccc aaa ggc aaa ctg ggg gag gta ttg     1410
Glu Gly Gly Lys Tyr Lys Tyr Pro Lys Gly Lys Leu Gly Glu Val Leu
420                 425                 430                 435

gcg ctc ctg gcc gcc cgg gag aac atg ggg tcc aac gag ggg tcg gag     1458
Ala Leu Leu Ala Ala Arg Glu Asn Met Gly Ser Asn Glu Gly Ser Glu
                440                 445                 450

gag gct tcg gac gaa cag tcc gag gag gag tcg gag gac acc gag agc     1506
Glu Ala Ser Asp Glu Gln Ser Glu Glu Glu Ser Glu Asp Thr Glu Ser
            455                 460                 465

gag gcg tcg gag ccg gag gac agg gca tcc agg aag ccc cgg gcc aag     1554
Glu Ala Ser Glu Pro Glu Asp Arg Ala Ser Arg Lys Pro Arg Ala Lys
        470                 475                 480

agg gcg cgc acg gcc ccc agg ggc ctg act ccg gcc ggc gcg cct ccc     1602
Arg Ala Arg Thr Ala Pro Arg Gly Leu Thr Pro Ala Gly Ala Pro Pro
    485                 490                 495

acc gct tcc ggg gcc cgc aaa acc cgc gcg ggc ggc cga ggc cga ggc     1650
Thr Ala Ser Gly Ala Arg Lys Thr Arg Ala Gly Gly Arg Gly Arg Gly
500                 505                 510                 515

cga ggc cgg ggc gtc act ccc gag aag aaa gcc ggg agc cgg ggc tcg     1698
Arg Gly Arg Gly Val Thr Pro Glu Lys Lys Ala Gly Ser Arg Gly Ser
                520                 525                 530

gcc cag gac gac gcc gca gga agc agg aag aag aag ggg agc gcg ggc     1746
Ala Gln Asp Asp Ala Ala Gly Ser Arg Lys Lys Lys Gly Ser Ala Gly
            535                 540                 545

tcc ggg gcc cat gcc agg gca ggc gag gcc aag ggc cag gcg ccc act     1794
Ser Gly Ala His Ala Arg Ala Gly Glu Ala Lys Gly Gln Ala Pro Thr
        550                 555                 560

gga tcc aag gcc gcg cgc ggg aag aag gcc cgt cgg ggc cgg agg ctg     1842
Gly Ser Lys Ala Ala Arg Gly Lys Lys Ala Arg Arg Gly Arg Arg Leu
    565                 570                 575
```

```
ccc cct aaa tgc cgc tag tggccc cccaagaagc cgcccaggct gcgagcaggc     1896
Pro Pro Lys Cys Arg  *
580                 585

cccgcagggc acccgcccgc ctgtggcccc cgccctcccc tcccctcttc ctgtcctccg   1956

cagacgcaat ctcctcgctt cacagcgcgc ccgggccgcg ttttgccagc gtcacgttcc   2016

cctctcgggc cctcgcaggc cgggggcgcc agcgatcccg acggaggaag cccggatggg   2076

aggaggaaag agaagtgggc gcccgaggca gcagcgcagg gccgagatgg ggacgcgcca   2136

agtggaccag gattgggggc ccgggttgcc cccggagggg gtgtgtgtgt ggacgccggg   2196

cacctgcaga ggcgagcagg gctcttcgtg gcgctctcgg ggcctgcgcc tggcaggtgc   2256

tgtaggccgc tgtcgcccct accccagtct gactgggccc tgggtctgtg gtggaggctc   2316

agtcaccagc cgcgcagcgc gtgtcagggc gcaactctca gccaggggag gccccagctc   2376

ccagccaggg gaagagatga ttccagaaag gaaagtctga gagatagaag gcggttggga   2436

aggggaggag gaggaaaggg gagaggaacg gtgggagaag ggaaagagga ggaggagggg   2496

gaggggggag cagagggaag acacatgcca gccctgccta ctggggcgcc cctgataaca   2556

aaggaaccag ccccaggcca aggccacctg cccctgacca caagttgaat ttgtcactca   2616

gactgcagtg tttcccaaca ttctaattat ttgcagaggt gttcaatttg gggtaattca   2676

cttaaaatcc agttttggtt cttctgggct gagtgggccc tggcccctcc cataggctgt   2736

ggctcccctg ggtgccccct ctccagtgga gctgacccac cgctcagcgc tggccttgca   2796

gcccttacta aaagacttga aagtccctgg gttcaccccc tgagtgaatt aaaggccaga   2856

ggggccccga agggcactgt gagggacaga ggctcacctg ggcagtgcag aagccggccg   2916

cgtgtccctc cttacagggg atgaaatgac ctggggagga aaccccagcc ctgccctgga   2976

ggttccagag taggcgggcc ggtgctgtga ggcttcacaa cctgctgtcc caagcacgct   3036

tgagttgtat gtgagtctgt gccgtgccgt gccgtatgct tcagctcctg caaccccggc   3096

tgagctcgat ttttacctaa atatcagtct ccacgggacc ccaccttcat tcatgccttc   3156

ttgtccctgg ggcaatgtgt gtgcttcctc gtcccaattt ccattccctg gcagtgagga   3216

gcccatcgtg ccaggggccc ctgccccact tgtccctggg aaggaatagg agggtttggg   3276

tgtgacctca cagtccagac cagactgtcc cagtcctatg tcagggacac ccagatgtag   3336

aagctgactg agacctgctg cagggcgtgg gtgctcccct ctgcttggag gctgtccctg   3396

gacagtgacc cacccactga ggaccaggct gggtgtacct tgagctgggc acagcagcct   3456
```

```
gtggtgttgc ctgtgggtgg ggagggcccc aggtgtgctt ctcccgtagc agtcctaggc   3516

ttctctccct gtgccctgtg tcacctggat cctccagtaa agtgaaattc agcactgtaa   3576

aaaaaaaaaa                                                          3586
```

<210> 20
<211> 3340
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (47)..(1894)

<400> 20

```
gatacagttc tgaccatcat tacgccaatc ttggcacgag gggaag   atg gcg gag      55
                                                     Met Ala Glu
                                                       1

tcc ggc ggt agc agc ggt ggt gct ggt ggc ggc ggc gct ttc ggc gcg      103
Ser Gly Gly Ser Ser Gly Gly Ala Gly Gly Gly Gly Ala Phe Gly Ala
      5                  10                  15

ggc ccg ggc ccc gag cgc ccg aac agc acg gcc gac aag aac ggg gcc      151
Gly Pro Gly Pro Glu Arg Pro Asn Ser Thr Ala Asp Lys Asn Gly Ala
 20                  25                  30                  35

ctc aag tgc acc ttc tcg gca ccc agc cac agc acc agc ctc ctg cag      199
Leu Lys Cys Thr Phe Ser Ala Pro Ser His Ser Thr Ser Leu Leu Gln
                 40                  45                  50

ggc ctg gcc acc ctc cgc gct cag ggc cag ctc ctc gat gtt gtg ctg      247
Gly Leu Ala Thr Leu Arg Ala Gln Gly Gln Leu Leu Asp Val Val Leu
             55                  60                  65

act att aac aga gag gcc ttt cct gca cac aag gtc gtc ctg gct gcc      295
Thr Ile Asn Arg Glu Ala Phe Pro Ala His Lys Val Val Leu Ala Ala
         70                  75                  80

tgc agc gac tac ttc agg gcc atg ttc acc ggc ggc atg cgg gag gca      343
Cys Ser Asp Tyr Phe Arg Ala Met Phe Thr Gly Gly Met Arg Glu Ala
     85                  90                  95

agc cag gac gtc atc gag ctg aag ggc gtg tcg gcc cgt ggc ctg cgg      391
Ser Gln Asp Val Ile Glu Leu Lys Gly Val Ser Ala Arg Gly Leu Arg
100                 105                 110                 115

cac atc atc gac ttc gcc tac agc gcc gag gtg aca ctg gac ctg gac      439
His Ile Ile Asp Phe Ala Tyr Ser Ala Glu Val Thr Leu Asp Leu Asp
                120                 125                 130

tgc gtg cag gac gtg ctg ggc gcg gcc gtg ttc ttg cag atg ctg ccc      487
Cys Val Gln Asp Val Leu Gly Ala Ala Val Phe Leu Gln Met Leu Pro
```

```
            135                 140                 145

gtg gtg gag ctg tgc gag gag ttc ctg aag gcg gcc atg agc gtg gag      535
Val Val Glu Leu Cys Glu Glu Phe Leu Lys Ala Ala Met Ser Val Glu
        150                 155                 160

acc tgc ctc aac atc ggc cag atg gcc acc acc ttc agc ctg gcc tcg      583
Thr Cys Leu Asn Ile Gly Gln Met Ala Thr Thr Phe Ser Leu Ala Ser
    165                 170                 175

ctg cga gag tcg gtg gat gcc ttc acc ttc cgg cac ttc ctg cag atc      631
Leu Arg Glu Ser Val Asp Ala Phe Thr Phe Arg His Phe Leu Gln Ile
180                 185                 190                 195

gcc gag gag gag gat ttc ctg cgc ctg cca ctg gag cgc ctg gtc ttc      679
Ala Glu Glu Glu Asp Phe Leu Arg Leu Pro Leu Glu Arg Leu Val Phe
                200                 205                 210

ttc ctg cag agc aac cgg ctg cag agc tgt gcc gag atc gac ctg ttc      727
Phe Leu Gln Ser Asn Arg Leu Gln Ser Cys Ala Glu Ile Asp Leu Phe
            215                 220                 225

cgc gcg gcc gtc cgc tgg ctg cag cat gac ccg gcc cgg cgg ccg cgc      775
Arg Ala Ala Val Arg Trp Leu Gln His Asp Pro Ala Arg Arg Pro Arg
        230                 235                 240

gcc agc cac gtg ctc tgc cac att cgc ttc ccg ctc atg cag tcg tcc      823
Ala Ser His Val Leu Cys His Ile Arg Phe Pro Leu Met Gln Ser Ser
    245                 250                 255

gag ctg gtg gac agc gtg cag acg ctg gac atc atg gtg gag gac gtg      871
Glu Leu Val Asp Ser Val Gln Thr Leu Asp Ile Met Val Glu Asp Val
260                 265                 270                 275

ctg tgc cgc cag tat ctg ctg gag gcc ttc aac tac cag gtg ctg ccc      919
Leu Cys Arg Gln Tyr Leu Leu Glu Ala Phe Asn Tyr Gln Val Leu Pro
                280                 285                 290

ttc cgg cag cac gag atg cag tct ccg cgc acc gcc gtg cgc tcg gat      967
Phe Arg Gln His Glu Met Gln Ser Pro Arg Thr Ala Val Arg Ser Asp
            295                 300                 305

gtg ccc tcg ctc gtc acc ttc ggc ggc acg ccc tac acc gac agc gac     1015
Val Pro Ser Leu Val Thr Phe Gly Gly Thr Pro Tyr Thr Asp Ser Asp
        310                 315                 320

cgc tcg gtc agc agc aag gtc tac cag ctg cct gag ccg gga gcc cgc     1063
Arg Ser Val Ser Ser Lys Val Tyr Gln Leu Pro Glu Pro Gly Ala Arg
    325                 330                 335

cac ttc cgc gag ctc acg gag atg gag gta ggc tgc agc cac acg tgc     1111
His Phe Arg Glu Leu Thr Glu Met Glu Val Gly Cys Ser His Thr Cys
340                 345                 350                 355

gtg gcc gtg ctg gac aat ttt gtg tac gtg gcc ggg ggg cag cac ctg     1159
Val Ala Val Leu Asp Asn Phe Val Tyr Val Ala Gly Gly Gln His Leu
                360                 365                 370
```

```
cag tac cgc agc ggc gag ggc gca gtg gac gcc tgc tac cgc tac gac     1207
Gln Tyr Arg Ser Gly Glu Gly Ala Val Asp Ala Cys Tyr Arg Tyr Asp
            375                 380                 385

ccc cac ctg aat cgc tgg ctg cgc ctg cag gcc atg cag gaa agc cgc     1255
Pro His Leu Asn Arg Trp Leu Arg Leu Gln Ala Met Gln Glu Ser Arg
        390                 395                 400

atc cag ttc cag ctg aac gtg ctg tgc ggc atg gtg tac gcc acg ggc     1303
Ile Gln Phe Gln Leu Asn Val Leu Cys Gly Met Val Tyr Ala Thr Gly
    405                 410                 415

ggc cgc aac cga gcc ggc agc ctg gcc tcc gtg gag cgg tac tgc ccc     1351
Gly Arg Asn Arg Ala Gly Ser Leu Ala Ser Val Glu Arg Tyr Cys Pro
420                 425                 430                 435

cgg cgc aat gag tgg ggc tac gcc tgc tcg ctg aag cgc cgt acc tgg     1399
Arg Arg Asn Glu Trp Gly Tyr Ala Cys Ser Leu Lys Arg Arg Thr Trp
                440                 445                 450

ggc cat gct ggg gcc gcc tca ggg ggc cgc ctc tac atc tcg ggt ggc     1447
Gly His Ala Gly Ala Ala Ser Gly Gly Arg Leu Tyr Ile Ser Gly Gly
            455                 460                 465

tac ggg atc tca gtg gag gac aag aag gcc ctg cac tgc tac gac ccc     1495
Tyr Gly Ile Ser Val Glu Asp Lys Lys Ala Leu His Cys Tyr Asp Pro
        470                 475                 480

gtg gcc gac cag tgg gag ttc aag gcg ccc atg agc gaa ccc cgc gtg     1543
Val Ala Asp Gln Trp Glu Phe Lys Ala Pro Met Ser Glu Pro Arg Val
    485                 490                 495

cta cac gcc atg gtg ggt gcc ggc ggc cgc atc tat gcc ctc ggg ggc     1591
Leu His Ala Met Val Gly Ala Gly Gly Arg Ile Tyr Ala Leu Gly Gly
500                 505                 510                 515

cgc atg gac cac gcg gac cgc tgc ttc gac gtg ctg gct gtg gag tac     1639
Arg Met Asp His Ala Asp Arg Cys Phe Asp Val Leu Ala Val Glu Tyr
                520                 525                 530

tat gtg ccg gag acg gac cag tgg acc agc gtg agc ccc atg cgg gcc     1687
Tyr Val Pro Glu Thr Asp Gln Trp Thr Ser Val Ser Pro Met Arg Ala
            535                 540                 545

ggc cag tca gag gcc ggc tgc tgc ctg ctg gag agg aag atc tac atc     1735
Gly Gln Ser Glu Ala Gly Cys Cys Leu Leu Glu Arg Lys Ile Tyr Ile
        550                 555                 560

gtc ggg ggc tac aac tgg cgt ctc aac aac gtc acg ggc atc gta cag     1783
Val Gly Gly Tyr Asn Trp Arg Leu Asn Asn Val Thr Gly Ile Val Gln
    565                 570                 575

gtg tac aac acg gac acc gac gag tgg gag cgg gac ctg cac ttc ccg     1831
Val Tyr Asn Thr Asp Thr Asp Glu Trp Glu Arg Asp Leu His Phe Pro
580                 585                 590                 595
```

```
gag tcc ttc gca ggc ata gcc tgc gcc ccc gtc ctg ctg ccc cgg gcc     1879
Glu Ser Phe Ala Gly Ile Ala Cys Ala Pro Val Leu Leu Pro Arg Ala
                600                 605                 610

ggg acc agg agg tag cccccaagac ccccgggacc ctggcctgac cgcatgttgt     1934
Gly Thr Arg Arg  *
            615

ctccaagtgg ggcttggcga atgcacgtct gcctgagaac cccagtgccc cccttcgccc   1994

gggctgccct tgaggggcct gctgcgttga taagcccccc tcccaggggt ccctccctcc   2054

ctccttccca aagcagatcc tggctgcgag tccatccgag ggagcctgcc ggcaaagcgt   2114

ctgacatgtg gtggcagcaa attcgtcccc ggggtggttt cctcgcctgg cccccgagtc   2174

cccacgggct ggcgggtgga atcccaggtc tccagggggt ccctgtgcag ctccatctca   2234

cttctctact gcctcccagc cccacggttt caggcattca gatgtgagct catcaacatt   2294

gaacccaaag tcggtggtat atgactcacc ctccttccaa gtctcctgcg cgcgtgtttt   2354

taaaataaac tcacccgaaa cgtccgtaac acacggacct ccaggagcag tgagaggtgg   2414

cttagaagcc cctggtttgg ggtgggtgga ggaggagggc acgtgtctgc ctcccctggg   2474

gtgccctcct tcccccatcc caagctgctg aggggaggcc ctggtcatgc ctcagttcct   2534

gccttcatct gctttccgga ggaaaaacca tatcaactcc tagaaacgct ccttaggggc   2594

ttgggacctt ccatttggca ctgagcatct tgtggggcct taactggctg agacatcccg   2654

gccctctaca tttgccctgt tggccaggca gtccccttcc cgcaattgga gggcgacgct   2714

aacttcagaa tcccatagtg gcgcttgccg caggtctggt ggggtgtctt ttttctcctc   2774

cctcctttcc acccctccgc gccctgccac tccctggcct gccctgtttt tgggtcaaca   2834

ttgctacgga gccagcagtg aggcctttcc cttcaagggc tctgtggtat ctctggccac   2894

atttgttcta atgtctgaac ctctaaatct tttctttttg attggtttta ctgtttttaa   2954

gaagccagca ctgctgtctc atagatggga tttgtactct tggggcaact taaagtgtct   3014

ctctcgctgc taacagacga ttgatgtctt gtctctgtga cccactcacc atgtaaagaa   3074

ttaacctcct atcttagcag acatcgtctc ctaatatttc cctttattta ataaaaatgt   3134

tatggtgaag agatggagcc ggcccagcac tgagcttgtg cggcttgggt ctgattggtc   3194

acagattcct cgtgtgtcct ccgcgtgtct gggggctcct ctcccccgcc tcagcctccc   3254

cccgcatccc cccaagaaag gaaaattatt tttcgtattg taaactttaa acatgaaaaa   3314

gctgtttttta atttaaaaaa aaaaaa                                       3340
```

<210> 21
<211> 2547
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (143)..(2275)

<400> 21
aaggatcctt aattaaatta atccccccc ccccccgcc cccgcggtgg cggcggcggc 60

ggcggtggct gccgtggcgg ctgagagtcc agagccggac gttccggccg cttcgggctg 120

gcggctggag agcgctcggg tc atg tct gcc cag ggg gac tgc gag ttc ctg 172
Met Ser Ala Gln Gly Asp Cys Glu Phe Leu
1 5 10

gtg cag cga gcc cgg gag ttg gtg ccg caa gac ctg tgg gca gcc aag 220
Val Gln Arg Ala Arg Glu Leu Val Pro Gln Asp Leu Trp Ala Ala Lys
15 20 25

gcg tgg ctg atc acg gcc cgc agc ctc tac ccg gca gac ttt aac atc 268
Ala Trp Leu Ile Thr Ala Arg Ser Leu Tyr Pro Ala Asp Phe Asn Ile
30 35 40

cag tat gag atg tac acc atc gag cgg aat gca gag cgg acc gcc acc 316
Gln Tyr Glu Met Tyr Thr Ile Glu Arg Asn Ala Glu Arg Thr Ala Thr
45 50 55

gcc ggg agg ctg ctg tac gac atg ttt gtg aat ttc cca gac cag ccg 364
Ala Gly Arg Leu Leu Tyr Asp Met Phe Val Asn Phe Pro Asp Gln Pro
60 65 70

gtg gtg tgg aga gaa atc agc att att aca tca gca tta agg aac gat 412
Val Val Trp Arg Glu Ile Ser Ile Ile Thr Ser Ala Leu Arg Asn Asp
75 80 85 90

tca cag gac aaa caa acc caa ttt tta aga agt tta ttt gaa act ctt 460
Ser Gln Asp Lys Gln Thr Gln Phe Leu Arg Ser Leu Phe Glu Thr Leu
95 100 105

cct ggt cgg gtc cag tgt gaa atg tta cta aag gtc acg gaa caa tgc 508
Pro Gly Arg Val Gln Cys Glu Met Leu Leu Lys Val Thr Glu Gln Cys
110 115 120

ttc aac acg tta gaa cga tca gaa atg ttg ctt cta ctt ttg agg cgc 556
Phe Asn Thr Leu Glu Arg Ser Glu Met Leu Leu Leu Leu Leu Arg Arg
125 130 135

ttc cct gaa acg gtg gtg cag cat ggg gtt ggc ctt ggg gag gca cta 604
Phe Pro Glu Thr Val Val Gln His Gly Val Gly Leu Gly Glu Ala Leu
140 145 150

tta gag gct gaa act att gaa gaa caa gaa tct cca gtg aac tgc ttt 652

```
Leu Glu Ala Glu Thr Ile Glu Glu Gln Glu Ser Pro Val Asn Cys Phe
155                 160                 165                 170

aga aaa tta ttt gtt tgt gat gtc ctt cct cta ata att aac aac cat      700
Arg Lys Leu Phe Val Cys Asp Val Leu Pro Leu Ile Ile Asn Asn His
                175                 180                 185

gat gtt cga tta cct gcc aat tta ttg tat aag tac ttg aac aaa gca      748
Asp Val Arg Leu Pro Ala Asn Leu Leu Tyr Lys Tyr Leu Asn Lys Ala
            190                 195                 200

gct gaa ttt tat atc aat tat gtc act agg tct act caa ata gaa aat      796
Ala Glu Phe Tyr Ile Asn Tyr Val Thr Arg Ser Thr Gln Ile Glu Asn
        205                 210                 215

cag cat caa ggc gcc cag gat aca tct gat tta atg tca cct agc aaa      844
Gln His Gln Gly Ala Gln Asp Thr Ser Asp Leu Met Ser Pro Ser Lys
    220                 225                 230

cgt agc tct cag aag tac ata ata gaa ggg ctg acg gaa aaa tca tcc      892
Arg Ser Ser Gln Lys Tyr Ile Ile Glu Gly Leu Thr Glu Lys Ser Ser
235                 240                 245                 250

cag atc gtg gac cct tgg gag agg ttg ttt aag att ttg aat gtt gtt      940
Gln Ile Val Asp Pro Trp Glu Arg Leu Phe Lys Ile Leu Asn Val Val
                255                 260                 265

gga atg aga tgt gaa tgg cag atg gat aaa gga aga cga agc tat gga      988
Gly Met Arg Cys Glu Trp Gln Met Asp Lys Gly Arg Arg Ser Tyr Gly
            270                 275                 280

gat att ttg cat aga atg aag gat ctc tgc aga tac atg aac aac ttt     1036
Asp Ile Leu His Arg Met Lys Asp Leu Cys Arg Tyr Met Asn Asn Phe
        285                 290                 295

gat agt gaa gca cat gca aaa tat aaa aac caa gtg gtg tat tcc acc     1084
Asp Ser Glu Ala His Ala Lys Tyr Lys Asn Gln Val Val Tyr Ser Thr
    300                 305                 310

atg ctg gtc ttc ttt aag aat gca ttc cag tat gtc aac agc ata cag     1132
Met Leu Val Phe Phe Lys Asn Ala Phe Gln Tyr Val Asn Ser Ile Gln
315                 320                 325                 330

cca tct ctc ttc caa ggt cct aat gcc ccg agc caa gtt cca ctg gtt     1180
Pro Ser Leu Phe Gln Gly Pro Asn Ala Pro Ser Gln Val Pro Leu Val
                335                 340                 345

ctt ctt gaa gat gta tcg aat gtg tat ggt gat gta gaa att gat cgt     1228
Leu Leu Glu Asp Val Ser Asn Val Tyr Gly Asp Val Glu Ile Asp Arg
            350                 355                 360

aat aaa cac atc cat aaa aag agg aaa cta gct gaa gga aga gaa aaa     1276
Asn Lys His Ile His Lys Lys Arg Lys Leu Ala Glu Gly Arg Glu Lys
        365                 370                 375

acc atg agt tca gac gat gaa gac tgt tcg gcg aaa gga aga aat cgt     1324
Thr Met Ser Ser Asp Asp Glu Asp Cys Ser Ala Lys Gly Arg Asn Arg
```

```
    380                 385                 390

cac att gta gtc aat aaa gcc gaa ctt gct aac tcc act gaa gtg tta     1372
His Ile Val Val Asn Lys Ala Glu Leu Ala Asn Ser Thr Glu Val Leu
395                 400                 405                 410

gaa agc ttt aaa ttg gcc agg gag agc tgg gag ttg ctc tat tcc cta     1420
Glu Ser Phe Lys Leu Ala Arg Glu Ser Trp Glu Leu Leu Tyr Ser Leu
                415                 420                 425

gaa ttc ctt gac aaa gaa ttt aca agg att tgc ttg gcc tgg aag acg     1468
Glu Phe Leu Asp Lys Glu Phe Thr Arg Ile Cys Leu Ala Trp Lys Thr
            430                 435                 440

gat act tgg ctt tgg tta aga atc ttc ctc act gat atg atc atc tat     1516
Asp Thr Trp Leu Trp Leu Arg Ile Phe Leu Thr Asp Met Ile Ile Tyr
        445                 450                 455

cag ggt caa tat aaa aag gcg ata gcc agc ctg cat cac tta gca gct     1564
Gln Gly Gln Tyr Lys Lys Ala Ile Ala Ser Leu His His Leu Ala Ala
    460                 465                 470

ctc cag gga tcc att tct cag cca cag atc aca ggg cag ggg acc ctg     1612
Leu Gln Gly Ser Ile Ser Gln Pro Gln Ile Thr Gly Gln Gly Thr Leu
475                 480                 485                 490

gag cat cag agg gcg ctc atc cag ctg gcg acg tgc cac ttt gcg cta     1660
Glu His Gln Arg Ala Leu Ile Gln Leu Ala Thr Cys His Phe Ala Leu
                495                 500                 505

ggg gag tac aga atg aca tgt gaa aaa gtc ctt gat ttg atg tgc tac     1708
Gly Glu Tyr Arg Met Thr Cys Glu Lys Val Leu Asp Leu Met Cys Tyr
            510                 515                 520

atg gta ctc ccc att caa gat gga ggc aaa tcc cag gag gaa ccc tcg     1756
Met Val Leu Pro Ile Gln Asp Gly Gly Lys Ser Gln Glu Glu Pro Ser
        525                 530                 535

aaa gta aag ccc aaa ttt aga aaa ggt tcg gat ctg aag ctc ctg cct     1804
Lys Val Lys Pro Lys Phe Arg Lys Gly Ser Asp Leu Lys Leu Leu Pro
    540                 545                 550

tgt acc agc aag gct atc atg cca tac tgc ctc cat tta atg tta gcc     1852
Cys Thr Ser Lys Ala Ile Met Pro Tyr Cys Leu His Leu Met Leu Ala
555                 560                 565                 570

tgt ttt aag ctt aga gct ttc aca gac aac aga gac gac atg gca ttg     1900
Cys Phe Lys Leu Arg Ala Phe Thr Asp Asn Arg Asp Asp Met Ala Leu
                575                 580                 585

ggg cat gtg att gtg ttg ctt cag caa gag tgg cca cgg ggc gag aat     1948
Gly His Val Ile Val Leu Leu Gln Gln Glu Trp Pro Arg Gly Glu Asn
            590                 595                 600

ctt ttc ctg aaa gct gtc aat aaa att tgc caa caa gga aat ttc caa     1996
Leu Phe Leu Lys Ala Val Asn Lys Ile Cys Gln Gln Gly Asn Phe Gln
        605                 610                 615
```

```
tat gag aat ttt ttc aat tac gtt aca aat att gat atg ctg gag gaa     2044
Tyr Glu Asn Phe Phe Asn Tyr Val Thr Asn Ile Asp Met Leu Glu Glu
    620                 625                 630

ttt gcc tac ttg aga act cag gaa ggt ggg aaa att cat ctg gaa tta     2092
Phe Ala Tyr Leu Arg Thr Gln Glu Gly Gly Lys Ile His Leu Glu Leu
635                 640                 645                 650

cta ccc aat caa gga atg ctg atc aag cac cac act gta act cga ggc     2140
Leu Pro Asn Gln Gly Met Leu Ile Lys His His Thr Val Thr Arg Gly
                655                 660                 665

atc acc aaa ggc gtg aag gag gac ttt cgc ctg gcc atg gag cgc cag     2188
Ile Thr Lys Gly Val Lys Glu Asp Phe Arg Leu Ala Met Glu Arg Gln
            670                 675                 680

gtc tcc cgc tgt gga gag aat ctg atg gtg gtt ctg cac agg ttc tgc     2236
Val Ser Arg Cys Gly Glu Asn Leu Met Val Val Leu His Arg Phe Cys
        685                 690                 695

att aat gag aag atc ttg ctc ctt cag act ctg acc tga gtggagacct     2285
Ile Asn Glu Lys Ile Leu Leu Leu Gln Thr Leu Thr  *
    700                 705                 710

ttccaccaga cacagctcgg gcctgtgtaa ttgtaggaga agacactcag cagtgattgc   2345

catggcacag agccgtggtc attgttgctg ttacaaagaa gaaaaccatc tgagttctaa   2405

ctccttggtt gcttaaaagt agttcccaag agtctgagaa gctatttcta tttttaagag   2465

tcatttttg taatttttgt aaaacaaaag taccaatctg ttttgtaaat aaaaatcatc   2525

ctaaaatttg aaaaaaaaaa aa                                           2547
```

<210> 22
<211> 2625
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (143)..(2353)

<400> 22

```
aaggatcctt aattaaatta atccccccc ccccccgcc cccgcggtgg cggcggcggc     60

ggcggtggct gccgtggcgg ctgagagtcc agagccggac gttccggccg cttcgggctg    120

gcggctggag agcgctcggg tc atg tct gcc cag ggg gac tgc gag ttc ctg     172
                        Met Ser Ala Gln Gly Asp Cys Glu Phe Leu
                          1               5                  10

gtg cag cga gcc cgg gag ttg gtg ccg caa gac ctg tgg gca gcc aag      220
```

```
Val Gln Arg Ala Arg Glu Leu Val Pro Gln Asp Leu Trp Ala Ala Lys
                15                  20                  25

gcg tgg ctg atc acg gcc cgc agc ctc tac ccg gca gac ttt aac atc      268
Ala Trp Leu Ile Thr Ala Arg Ser Leu Tyr Pro Ala Asp Phe Asn Ile
            30                  35                  40

cag tat gag atg tac acc atc gag cgg aat gca gag cgg acc gcc acc      316
Gln Tyr Glu Met Tyr Thr Ile Glu Arg Asn Ala Glu Arg Thr Ala Thr
        45                  50                  55

gcc ggg agg ctg ctg tac gac atg ttt gtg aat ttc cca gac cag ccg      364
Ala Gly Arg Leu Leu Tyr Asp Met Phe Val Asn Phe Pro Asp Gln Pro
    60                  65                  70

gtg gtg tgg aga gaa atc agc att att aca tca gca tta agg aac gat      412
Val Val Trp Arg Glu Ile Ser Ile Ile Thr Ser Ala Leu Arg Asn Asp
75                  80                  85                  90

tca cag gac aaa caa acc caa ttt tta aga agt tta ttt gaa act ctt      460
Ser Gln Asp Lys Gln Thr Gln Phe Leu Arg Ser Leu Phe Glu Thr Leu
                95                  100                 105

cct ggt cgg gtc cag tgt gaa atg tta cta aag gtc acg gaa caa tgc      508
Pro Gly Arg Val Gln Cys Glu Met Leu Leu Lys Val Thr Glu Gln Cys
            110                 115                 120

ttc aac acg tta gaa cga tca gaa atg ttg ctt cta ctt ttg agg cgc      556
Phe Asn Thr Leu Glu Arg Ser Glu Met Leu Leu Leu Leu Leu Arg Arg
        125                 130                 135

ttc cct gaa acg gtg gtg cag cat ggg gtt ggc ctt ggg gag gca cta      604
Phe Pro Glu Thr Val Val Gln His Gly Val Gly Leu Gly Glu Ala Leu
    140                 145                 150

tta gag gct gaa act att gaa gaa caa gaa tct cca gtg aac tgc ttt      652
Leu Glu Ala Glu Thr Ile Glu Glu Gln Glu Ser Pro Val Asn Cys Phe
155                 160                 165                 170

aga aaa tta ttt gtt tgt gat gtc ctt cct cta ata att aac aac cat      700
Arg Lys Leu Phe Val Cys Asp Val Leu Pro Leu Ile Ile Asn Asn His
                175                 180                 185

gat gtt cga tta cct gcc aat tta ttg tat aag tac ttg aac aaa gca      748
Asp Val Arg Leu Pro Ala Asn Leu Leu Tyr Lys Tyr Leu Asn Lys Ala
            190                 195                 200

gct gaa ttt tat atc aat tat gtc act agg tct act caa ata gaa aat      796
Ala Glu Phe Tyr Ile Asn Tyr Val Thr Arg Ser Thr Gln Ile Glu Asn
        205                 210                 215

cag cat caa ggc gcc cag gat aca tct gat tta atg tca cct agc aaa      844
Gln His Gln Gly Ala Gln Asp Thr Ser Asp Leu Met Ser Pro Ser Lys
    220                 225                 230

cgt agc tct cag aag tac ata ata gaa ggg ctg acg gaa aaa tca tcc      892
Arg Ser Ser Gln Lys Tyr Ile Ile Glu Gly Leu Thr Glu Lys Ser Ser
```

```
235                 240                 245                 250

cag atc gtg gac cct tgg gag agg ttg ttt aag att ttg aat gtt gtt      940
Gln Ile Val Asp Pro Trp Glu Arg Leu Phe Lys Ile Leu Asn Val Val
                255                 260                 265

gga atg aga tgt gaa tgg cag atg gat aaa gga aga cga agc tat gga      988
Gly Met Arg Cys Glu Trp Gln Met Asp Lys Gly Arg Arg Ser Tyr Gly
            270                 275                 280

gat att ttg cat aga atg aag gat ctc tgc aga tac atg aac aac ttt     1036
Asp Ile Leu His Arg Met Lys Asp Leu Cys Arg Tyr Met Asn Asn Phe
        285                 290                 295

gat agt gaa gca cat gca aaa tat aaa aac caa gtg gtg tat tcc acc     1084
Asp Ser Glu Ala His Ala Lys Tyr Lys Asn Gln Val Val Tyr Ser Thr
    300                 305                 310

atg ctg gtc ttc ttt aag aat gca ttc cag tat gtc aac agc ata cag     1132
Met Leu Val Phe Phe Lys Asn Ala Phe Gln Tyr Val Asn Ser Ile Gln
315                 320                 325                 330

cca tct ctc ttc caa ggt cct aat gcc ccg agc caa gtt cca ctg gtt     1180
Pro Ser Leu Phe Gln Gly Pro Asn Ala Pro Ser Gln Val Pro Leu Val
                335                 340                 345

ctt ctt gaa gat gta tcg aat gtg tat ggt gat gta gaa att gat cgt     1228
Leu Leu Glu Asp Val Ser Asn Val Tyr Gly Asp Val Glu Ile Asp Arg
            350                 355                 360

aat aaa cac atc cat aaa aag agg aaa cta gct gaa gga aga gaa aaa     1276
Asn Lys His Ile His Lys Lys Arg Lys Leu Ala Glu Gly Arg Glu Lys
        365                 370                 375

acc atg agt tca gac gat gaa gac tgt tcg gcg aaa gga aga aat cgt     1324
Thr Met Ser Ser Asp Asp Glu Asp Cys Ser Ala Lys Gly Arg Asn Arg
    380                 385                 390

cac att gta gtc aat aaa gcc gaa ctt gct aac tcc act gaa gtg tta     1372
His Ile Val Val Asn Lys Ala Glu Leu Ala Asn Ser Thr Glu Val Leu
395                 400                 405                 410

gaa agc ttt aaa ttg gcc agg gag agc tgg gag ttg ctc tat tcc cta     1420
Glu Ser Phe Lys Leu Ala Arg Glu Ser Trp Glu Leu Leu Tyr Ser Leu
                415                 420                 425

gaa ttc ctt gac aaa gaa ttt aca agg att tgc ttg gcc tgg aag acg     1468
Glu Phe Leu Asp Lys Glu Phe Thr Arg Ile Cys Leu Ala Trp Lys Thr
            430                 435                 440

gat act tgg ctt tgg tta aga atc ttc ctc act gat atg atc atc tat     1516
Asp Thr Trp Leu Trp Leu Arg Ile Phe Leu Thr Asp Met Ile Ile Tyr
        445                 450                 455

cag ggt caa tat aaa aag gcg ata gcc agc ctg cat cac tta gca gct     1564
Gln Gly Gln Tyr Lys Lys Ala Ile Ala Ser Leu His His Leu Ala Ala
    460                 465                 470
```

```
ctc cag gga tcc att tct cag cca cag atc aca ggg cag ggg acc ctg      1612
Leu Gln Gly Ser Ile Ser Gln Pro Gln Ile Thr Gly Gln Gly Thr Leu
475                 480                 485                 490

gag cat cag agg gcg ctc atc cag ctg gcg acg tgc cac ttt gcg cta      1660
Glu His Gln Arg Ala Leu Ile Gln Leu Ala Thr Cys His Phe Ala Leu
                495                 500                 505

ggg gag tac aga atg aca tgt gaa aaa gtc ctt gat ttg atg tgc tac      1708
Gly Glu Tyr Arg Met Thr Cys Glu Lys Val Leu Asp Leu Met Cys Tyr
            510                 515                 520

atg gta ctc ccc att caa gat gga ggc aaa tcc cag gag gaa ccc tcg      1756
Met Val Leu Pro Ile Gln Asp Gly Gly Lys Ser Gln Glu Glu Pro Ser
        525                 530                 535

aaa gta aag ccc aaa ttt aga aaa ggt tcg gat ctg aag ctc ctg cct      1804
Lys Val Lys Pro Lys Phe Arg Lys Gly Ser Asp Leu Lys Leu Leu Pro
    540                 545                 550

tgt acc agc aag gct atc atg cca tac tgc ctc cat tta atg tta gcc      1852
Cys Thr Ser Lys Ala Ile Met Pro Tyr Cys Leu His Leu Met Leu Ala
555                 560                 565                 570

tgt ttt aag ctt aga gct ttc aca gac aac aga gac gac atg gca ttg      1900
Cys Phe Lys Leu Arg Ala Phe Thr Asp Asn Arg Asp Asp Met Ala Leu
                575                 580                 585

ggg cat gtg att gtg ttg ctt cag caa gag tgg cca cgg ggc gag aat      1948
Gly His Val Ile Val Leu Leu Gln Gln Glu Trp Pro Arg Gly Glu Asn
            590                 595                 600

ctt ttc ctg aaa gct gtc aat aaa att tgc caa caa gga aat ttc caa      1996
Leu Phe Leu Lys Ala Val Asn Lys Ile Cys Gln Gln Gly Asn Phe Gln
        605                 610                 615

tat gag aat ttt ttc aat tac gtt aca aat att gat atg ctg gag gaa      2044
Tyr Glu Asn Phe Phe Asn Tyr Val Thr Asn Ile Asp Met Leu Glu Glu
    620                 625                 630

ttt gcc tac ttg aga act cag gaa ggt ggg aaa att cat ctg gaa tta      2092
Phe Ala Tyr Leu Arg Thr Gln Glu Gly Gly Lys Ile His Leu Glu Leu
635                 640                 645                 650

cta ccc aat caa gga atg ctg atc aag cct tct agc cct ccc atg ggg      2140
Leu Pro Asn Gln Gly Met Leu Ile Lys Pro Ser Ser Pro Pro Met Gly
                655                 660                 665

tta ctg cag cag gaa ttc tta cct gtg ctt cag ccc agc ata cag act      2188
Leu Leu Gln Gln Glu Phe Leu Pro Val Leu Gln Pro Ser Ile Gln Thr
            670                 675                 680

gct gac agg cac cac act gta act cga ggc atc acc aaa ggc gtg aag      2236
Ala Asp Arg His His Thr Val Thr Arg Gly Ile Thr Lys Gly Val Lys
        685                 690                 695
```

```
gag gac ttt cgc ctg gcc atg gag cgc cag gtc tcc cgc tgt gga gag     2284
Glu Asp Phe Arg Leu Ala Met Glu Arg Gln Val Ser Arg Cys Gly Glu
    700                 705                 710

aat ctg atg gtg gtt ctg cac agg ttc tgc att aat gag aag atc ttg     2332
Asn Leu Met Val Val Leu His Arg Phe Cys Ile Asn Glu Lys Ile Leu
715                 720                 725                 730

ctc ctt cag act ctg acc tga gt ggagaccttt ccaccagaca cagctcgggc     2385
Leu Leu Gln Thr Leu Thr  *
                735

ctgtgtaatt gtaggagaag acactcagca gtgattgcca tggcacagag ccgtggtcat   2445

tgttgctgtt acaaagaaga aaaccatctg agttctaact ccttggttgc ttaaaagtag   2505

ttcccaagag tctgagaagc tatttctatt tttaagagtc atttttttgta atttttgtaa   2565

aacaaaagta ccaatctgtt ttgtaaataa aaatcatcct aaaatttgaa aaaaaaaaaa   2625
```

<210> 23
<211> 6288
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (115)..(5253)

<400> 23

```
aagggagccc cgctcagcgc ggggagcgcc cggccccctc cccgccccat gcgcccgcgg     60

ctctgaagcc tgagcggggc cgggggctgg gcgggggccgg ggcccgccgt aggc  atg     117
                                                             Met
                                                               1

gcg tcc ggg agc cgg tgg cgg ccg aca ccg ccg ccg ctg ctg ttg ctg      165
Ala Ser Gly Ser Arg Trp Arg Pro Thr Pro Pro Pro Leu Leu Leu Leu
              5                  10                  15

ctg ctg ctg gcg ctg gcg gcg cgc gcg gac ggc ctg gag ttc ggc ggc      213
Leu Leu Leu Ala Leu Ala Ala Arg Ala Asp Gly Leu Glu Phe Gly Gly
         20                  25                  30

ggc ccc ggg cag tgg gct cgc tac gcg cgc tgg gcg ggc gcg gcg agc      261
Gly Pro Gly Gln Trp Ala Arg Tyr Ala Arg Trp Ala Gly Ala Ala Ser
     35                  40                  45

agc ggc gag ctc agc ttc agc ctg cgc acc aac gcc acg cgc gcg ctg      309
Ser Gly Glu Leu Ser Phe Ser Leu Arg Thr Asn Ala Thr Arg Ala Leu
 50                  55                  60                  65

ctg ctc tac ctg gac gac ggc ggc gac tgc gac ttc ctg gag ctg ctg      357
Leu Leu Tyr Leu Asp Asp Gly Gly Asp Cys Asp Phe Leu Glu Leu Leu
```

```
                    70                  75                  80

ctg gtg gac ggc cgc ctg cgg ctg cgc ttc acg ctt tcg tgc gcc gag      405
Leu Val Asp Gly Arg Leu Arg Leu Arg Phe Thr Leu Ser Cys Ala Glu
                85                  90                  95

ccg gcc acg ctg cag ctg gac acg ccg gtg gcc gac gac cgc tgg cac      453
Pro Ala Thr Leu Gln Leu Asp Thr Pro Val Ala Asp Asp Arg Trp His
        100                 105                 110

atg gtg ctg ctg acc cgc gac gcg cgc cgc acg gcg ctg gcg gtg gac      501
Met Val Leu Leu Thr Arg Asp Ala Arg Arg Thr Ala Leu Ala Val Asp
    115                 120                 125

ggc gag gcc cgc gcc gcc gag gtg cgc tcc aag cgg cgc gag atg cag      549
Gly Glu Ala Arg Ala Ala Glu Val Arg Ser Lys Arg Arg Glu Met Gln
130                 135                 140                 145

gtg gcc agc gac ctg ttc gtg ggc ggc atc ccg ccc gac gtg cgc ctc      597
Val Ala Ser Asp Leu Phe Val Gly Gly Ile Pro Pro Asp Val Arg Leu
                150                 155                 160

tcg gcg ctt acg ctg agc acc gtc aag tac gag ccg ccc ttc cgc ggt      645
Ser Ala Leu Thr Leu Ser Thr Val Lys Tyr Glu Pro Pro Phe Arg Gly
            165                 170                 175

ctc ttg gcc aac ctg aag ctg ggc gag cgg ccc ccc gcg ctg ctg ggc      693
Leu Leu Ala Asn Leu Lys Leu Gly Glu Arg Pro Pro Ala Leu Leu Gly
        180                 185                 190

agc cag ggc ctg cgc ggc gcc acc gcc gac ccg ctg tgc gcg ccc gcg      741
Ser Gln Gly Leu Arg Gly Ala Thr Ala Asp Pro Leu Cys Ala Pro Ala
    195                 200                 205

cgc aac ccc tgc gcc aac ggc ggc ctc tgc acc gtg ctg gcc ccc ggc      789
Arg Asn Pro Cys Ala Asn Gly Gly Leu Cys Thr Val Leu Ala Pro Gly
210                 215                 220                 225

gag gtg ggc tgc gac tgc agc cac acg ggc ttc ggc ggc aag ttc tgc      837
Glu Val Gly Cys Asp Cys Ser His Thr Gly Phe Gly Gly Lys Phe Cys
                230                 235                 240

agc gaa gag gag cac ccc atg gaa ggt ccg gct cac ctg acg tta aac      885
Ser Glu Glu Glu His Pro Met Glu Gly Pro Ala His Leu Thr Leu Asn
            245                 250                 255

agc gaa gta ggg tcc tta ctg ttc tcc gag ggg ggg gcc ggg aga gga      933
Ser Glu Val Gly Ser Leu Leu Phe Ser Glu Gly Gly Ala Gly Arg Gly
        260                 265                 270

gga gcc ggc gat gtg cac cag cca aca aaa ggc aag gag gag ttt gtg      981
Gly Ala Gly Asp Val His Gln Pro Thr Lys Gly Lys Glu Glu Phe Val
    275                 280                 285

gcg acc ttc aaa ggc aat gag ttc ttc tgc tac gac ctg tca cac aac     1029
Ala Thr Phe Lys Gly Asn Glu Phe Phe Cys Tyr Asp Leu Ser His Asn
290                 295                 300                 305
```

```
ccc atc cag agc agc act gat gag atc aca ctg gcc ttc cgc acc ctg      1077
Pro Ile Gln Ser Ser Thr Asp Glu Ile Thr Leu Ala Phe Arg Thr Leu
                310                 315                 320

caa cgc aac ggc ctg atg ctg cat aca ggc aag tcg gcc gac tac gtc      1125
Gln Arg Asn Gly Leu Met Leu His Thr Gly Lys Ser Ala Asp Tyr Val
            325                 330                 335

aac ctg tcc ctc aag tct ggg gct gtc tgg ctg gtc atc aac cta ggc      1173
Asn Leu Ser Leu Lys Ser Gly Ala Val Trp Leu Val Ile Asn Leu Gly
        340                 345                 350

tca ggt gcc ttc gag gcc ctt gtg gaa ccc gtc aat ggc aag ttc aac      1221
Ser Gly Ala Phe Glu Ala Leu Val Glu Pro Val Asn Gly Lys Phe Asn
    355                 360                 365

gac aac gcc tgg cac gac gtc cgg gtc acc cga aac ctg cgc cag cac      1269
Asp Asn Ala Trp His Asp Val Arg Val Thr Arg Asn Leu Arg Gln His
370                 375                 380                 385

gca ggg att gga cac gct atg gta aac aaa ctg cat tat ctg gtg acc      1317
Ala Gly Ile Gly His Ala Met Val Asn Lys Leu His Tyr Leu Val Thr
                390                 395                 400

atc tcg gtg gac ggg atc ctg acc acc aca ggc tac acg cag gag gat      1365
Ile Ser Val Asp Gly Ile Leu Thr Thr Thr Gly Tyr Thr Gln Glu Asp
            405                 410                 415

tac acc atg ctg ggc tct gat gac ttc ttc tac att ggg ggc agc ccc      1413
Tyr Thr Met Leu Gly Ser Asp Asp Phe Phe Tyr Ile Gly Gly Ser Pro
        420                 425                 430

aac aca gct gac ctg ccg ggc tcg ccc gtc agc aac aac ttc atg ggc      1461
Asn Thr Ala Asp Leu Pro Gly Ser Pro Val Ser Asn Asn Phe Met Gly
    435                 440                 445

tgc ctc aag gac gtg gtc tat aag aac aat gac ttc aaa ttg gaa cta      1509
Cys Leu Lys Asp Val Val Tyr Lys Asn Asn Asp Phe Lys Leu Glu Leu
450                 455                 460                 465

tcc cgc ctg gca aag gaa ggg gac ccc aag atg aag ctg cag ggg gac      1557
Ser Arg Leu Ala Lys Glu Gly Asp Pro Lys Met Lys Leu Gln Gly Asp
                470                 475                 480

ttg tca ttc cgc tgt gag gat gtg gct gcc ctg gac cct gtg acc ttt      1605
Leu Ser Phe Arg Cys Glu Asp Val Ala Ala Leu Asp Pro Val Thr Phe
            485                 490                 495

gag agt ccc gag gcc ttt gtg gcg ctg ccc cgc tgg agc gct aag cgc      1653
Glu Ser Pro Glu Ala Phe Val Ala Leu Pro Arg Trp Ser Ala Lys Arg
        500                 505                 510

act ggc tcc atc tcc cta gac ttc cgc acc acc gag ccc aat ggg ctg      1701
Thr Gly Ser Ile Ser Leu Asp Phe Arg Thr Thr Glu Pro Asn Gly Leu
    515                 520                 525
```

```
ctg ctc ttc agc cag ggc cgg cgg gct ggg ggt gga gct ggc agc cac      1749
Leu Leu Phe Ser Gln Gly Arg Arg Ala Gly Gly Gly Ala Gly Ser His
530                 535                 540                 545

agc tct gct cag cgg gcc gac tac ttt gcc atg gag cta ttg gac ggc      1797
Ser Ser Ala Gln Arg Ala Asp Tyr Phe Ala Met Glu Leu Leu Asp Gly
                550                 555                 560

cac ctc tat ctt ctg ctg gac atg gga tct ggg ggc atc aag ctg cgg      1845
His Leu Tyr Leu Leu Leu Asp Met Gly Ser Gly Gly Ile Lys Leu Arg
            565                 570                 575

gca tcc agc cgc aag gtc aat gat ggc gag tgg tgt cac gtg gac ttc      1893
Ala Ser Ser Arg Lys Val Asn Asp Gly Glu Trp Cys His Val Asp Phe
        580                 585                 590

cag agg gat ggg cga aaa ggc tcc atc tca gtg aat agt cgc agc acg      1941
Gln Arg Asp Gly Arg Lys Gly Ser Ile Ser Val Asn Ser Arg Ser Thr
    595                 600                 605

ccg ttc ttg gcc act gga gac agc gag att ctg gac ctg gag agt gag      1989
Pro Phe Leu Ala Thr Gly Asp Ser Glu Ile Leu Asp Leu Glu Ser Glu
610                 615                 620                 625

ctg tac ctg ggc ggt ctc cct gag ggg ggc cgg gtg gac ctg ccc ctg      2037
Leu Tyr Leu Gly Gly Leu Pro Glu Gly Gly Arg Val Asp Leu Pro Leu
                630                 635                 640

ccc cca gag gtg tgg aca gca gca ctc cgg gca ggc tac gtg ggc tgt      2085
Pro Pro Glu Val Trp Thr Ala Ala Leu Arg Ala Gly Tyr Val Gly Cys
            645                 650                 655

gtg cgg gac ctc ttc ata gat ggg cgt agc cga gac ctc cgg ggc ctg      2133
Val Arg Asp Leu Phe Ile Asp Gly Arg Ser Arg Asp Leu Arg Gly Leu
        660                 665                 670

gct gag gct cag ggg gct gtg ggc gtt gcc ccc ttt tgc tcc cgg gag      2181
Ala Glu Ala Gln Gly Ala Val Gly Val Ala Pro Phe Cys Ser Arg Glu
    675                 680                 685

acg ctg aag cag tgt gca tct gcc ccc tgt cgc aat ggg ggc gtc tgt      2229
Thr Leu Lys Gln Cys Ala Ser Ala Pro Cys Arg Asn Gly Gly Val Cys
690                 695                 700                 705

cga gaa ggc tgg aac cgc ttc atc tgt gac tgc atc ggg acc ggc ttt      2277
Arg Glu Gly Trp Asn Arg Phe Ile Cys Asp Cys Ile Gly Thr Gly Phe
                710                 715                 720

ctt ggg cgg gtc tgt gag aga gag gcc acg gtc ctg agc tac gat ggc      2325
Leu Gly Arg Val Cys Glu Arg Glu Ala Thr Val Leu Ser Tyr Asp Gly
            725                 730                 735

tcc atg tac atg aag atc atg ctg cct aac gcc atg cac acg gag gca      2373
Ser Met Tyr Met Lys Ile Met Leu Pro Asn Ala Met His Thr Glu Ala
        740                 745                 750

gag gat gtg tcc ctg cgt ttc atg tcc cag cgg gcc tac gga ctc atg      2421
```

```
Glu Asp Val Ser Leu Arg Phe Met Ser Gln Arg Ala Tyr Gly Leu Met
    755                 760                 765

atg gcc acc act tcc agg gag tct gcc gac acc cta cgc ctg gag ctg     2469
Met Ala Thr Thr Ser Arg Glu Ser Ala Asp Thr Leu Arg Leu Glu Leu
770                 775                 780                 785

gat ggg ggg cag atg aag ctc act gtc aac ctc gac tgc ctg cgc gtc     2517
Asp Gly Gly Gln Met Lys Leu Thr Val Asn Leu Asp Cys Leu Arg Val
                790                 795                 800

ggc tgc gca ccc agt aaa ggc ccc gaa acg ctg ttt gcg ggg cac aag     2565
Gly Cys Ala Pro Ser Lys Gly Pro Glu Thr Leu Phe Ala Gly His Lys
            805                 810                 815

ctc aat gac aat gag tgg cac acg gtg agg gtg gtc cgg cgt ggc aag     2613
Leu Asn Asp Asn Glu Trp His Thr Val Arg Val Val Arg Arg Gly Lys
        820                 825                 830

agc ctg cag ctg tct gtg gac aac gtg act gtg gag gga cag atg gca     2661
Ser Leu Gln Leu Ser Val Asp Asn Val Thr Val Glu Gly Gln Met Ala
    835                 840                 845

gga gcc cat atg cgg ctg gag ttc cac aac att gag acg ggc atc atg     2709
Gly Ala His Met Arg Leu Glu Phe His Asn Ile Glu Thr Gly Ile Met
850                 855                 860                 865

acg gag cgg cgg ttt atc tcc gtg gtg ccc tcc aac ttc atc ggg cat     2757
Thr Glu Arg Arg Phe Ile Ser Val Val Pro Ser Asn Phe Ile Gly His
                870                 875                 880

ctg agt ggg ctc gtg ttc aat ggc cag ccc tac atg gac cag tgc aag     2805
Leu Ser Gly Leu Val Phe Asn Gly Gln Pro Tyr Met Asp Gln Cys Lys
            885                 890                 895

gat ggt gac atc acc tac tgt gag ctc aat gct cgc ttt ggc ctg cgt     2853
Asp Gly Asp Ile Thr Tyr Cys Glu Leu Asn Ala Arg Phe Gly Leu Arg
        900                 905                 910

gcc att gtg gcc gat ccc gtc acc ttc aag agt cgc agc agc tac ctg     2901
Ala Ile Val Ala Asp Pro Val Thr Phe Lys Ser Arg Ser Ser Tyr Leu
    915                 920                 925

gca ctc gcc acg ctc caa gcc tat gct tcc atg cac ctc ttc ttc cag     2949
Ala Leu Ala Thr Leu Gln Ala Tyr Ala Ser Met His Leu Phe Phe Gln
930                 935                 940                 945

ttc aag acc acg gcc cct gat ggg ctt ctt ctg ttc aac tcg ggc aac     2997
Phe Lys Thr Thr Ala Pro Asp Gly Leu Leu Leu Phe Asn Ser Gly Asn
                950                 955                 960

ggc aat gac ttc att gtc atc gag ctg gtc aag ggg tac atc cac tac     3045
Gly Asn Asp Phe Ile Val Ile Glu Leu Val Lys Gly Tyr Ile His Tyr
            965                 970                 975

gtg ttt gac ctg ggg aat ggc ccg tcc ttg atg aag ggg aac tca gac     3093
Val Phe Asp Leu Gly Asn Gly Pro Ser Leu Met Lys Gly Asn Ser Asp
```

```
        980                 985                 990

aaa cca gtc aat gac aac cag tgg cac aac gtg gtg gtg tcc agg gac     3141
Lys Pro Val Asn Asp Asn Gln Trp His Asn Val Val Val Ser Arg Asp
    995                1000                1005

cca ggc aac gtg cac acg ctc aag att gac tcc cgc act gtc acg cag     3189
Pro Gly Asn Val His Thr Leu Lys Ile Asp Ser Arg Thr Val Thr Gln
1010                1015                1020                1025

cac tcc aat ggc gcc cga aac ctc gat ctc aaa ggg gag ttg tac att     3237
His Ser Asn Gly Ala Arg Asn Leu Asp Leu Lys Gly Glu Leu Tyr Ile
                1030                1035                1040

ggc ggt ctg agc aag aat atg ttc agc aac ctg ccc aag ctg gtg gcc     3285
Gly Gly Leu Ser Lys Asn Met Phe Ser Asn Leu Pro Lys Leu Val Ala
            1045                1050                1055

tcc cgg gat ggc ttt cag ggc tgc ctg gcc tca gtg gac ctc aac gga     3333
Ser Arg Asp Gly Phe Gln Gly Cys Leu Ala Ser Val Asp Leu Asn Gly
        1060                1065                1070

cgt ctc cca gac ctc atc gcc gac gcc ctg cac cgc att ggg cag gtg     3381
Arg Leu Pro Asp Leu Ile Ala Asp Ala Leu His Arg Ile Gly Gln Val
    1075                1080                1085

gag agg ggc tgt gat ggc ccc agc acc acc tgc act gaa gag tcc tgt     3429
Glu Arg Gly Cys Asp Gly Pro Ser Thr Thr Cys Thr Glu Glu Ser Cys
1090                1095                1100                1105

gcc aac cag ggc gtc tgc ttg cag cag tgg gat ggc ttc acc tgc gac     3477
Ala Asn Gln Gly Val Cys Leu Gln Gln Trp Asp Gly Phe Thr Cys Asp
                1110                1115                1120

tgc acc atg act tcc tat gga ggc cct gtc tgc aat gat ccc ggg acc     3525
Cys Thr Met Thr Ser Tyr Gly Gly Pro Val Cys Asn Asp Pro Gly Thr
            1125                1130                1135

aca tac atc ttt ggg aag ggg gga gcg ctc atc acc tac acg tgg ccc     3573
Thr Tyr Ile Phe Gly Lys Gly Gly Ala Leu Ile Thr Tyr Thr Trp Pro
        1140                1145                1150

ccc aat gac agg ccc agc acg agg atg gat cgc ctg gcc gtg ggc ttc     3621
Pro Asn Asp Arg Pro Ser Thr Arg Met Asp Arg Leu Ala Val Gly Phe
    1155                1160                1165

agc acc cac cag cgg agc gct gtg ctg gtg cgg gtg gac agc gcc tcc     3669
Ser Thr His Gln Arg Ser Ala Val Leu Val Arg Val Asp Ser Ala Ser
1170                1175                1180                1185

ggc ctt gga gac tac ctg cag ctg cac atc gac cag ggc acc gtg ggg     3717
Gly Leu Gly Asp Tyr Leu Gln Leu His Ile Asp Gln Gly Thr Val Gly
                1190                1195                1200

gtg atc ttt aac gtg ggc acg gac gac att acc atc gac gag ccc aac     3765
Val Ile Phe Asn Val Gly Thr Asp Asp Ile Thr Ile Asp Glu Pro Asn
            1205                1210                1215
```

```
gcc ata gta agc gac ggc aaa tac cac gtg gtg cgc ttc act cga agc     3813
Ala Ile Val Ser Asp Gly Lys Tyr His Val Val Arg Phe Thr Arg Ser
       1220                1225                1230

ggc ggc aac gcc acc ctg cag gtg gac agc tgg ccg gtc aac gag cgg     3861
Gly Gly Asn Ala Thr Leu Gln Val Asp Ser Trp Pro Val Asn Glu Arg
   1235                1240                1245

tac ccg gca gga aac ttt gat aac gag cgc ctg gcg att gct aga cag     3909
Tyr Pro Ala Gly Asn Phe Asp Asn Glu Arg Leu Ala Ile Ala Arg Gln
1250                1255                1260                1265

aga atc ccc tac cgg ctt ggt cga gta gta gat gag tgg ctg ctc gac     3957
Arg Ile Pro Tyr Arg Leu Gly Arg Val Val Asp Glu Trp Leu Leu Asp
               1270                1275                1280

aaa ggc cgc cag ctg acc atc ttc aac agc cag gct gcc atc aag atc     4005
Lys Gly Arg Gln Leu Thr Ile Phe Asn Ser Gln Ala Ala Ile Lys Ile
           1285                1290                1295

ggg ggc cgg gat cag ggc cgc ccc ttc cag ggc cag gtg tcc ggc ctc     4053
Gly Gly Arg Asp Gln Gly Arg Pro Phe Gln Gly Gln Val Ser Gly Leu
       1300                1305                1310

tac tac aat ggg ctc aag gtg ctg gcg ctg gcc gcc gag agc gac ccc     4101
Tyr Tyr Asn Gly Leu Lys Val Leu Ala Leu Ala Ala Glu Ser Asp Pro
   1315                1320                1325

aat gtg cgg act gag ggt cac ctg cgc ctg gtg ggg gag ggg ccg tcc     4149
Asn Val Arg Thr Glu Gly His Leu Arg Leu Val Gly Glu Gly Pro Ser
1330                1335                1340                1345

gtg ctg ctc agt gcg gag acc acg gcc acc acc ctg ctg gct gac atg     4197
Val Leu Leu Ser Ala Glu Thr Thr Ala Thr Thr Leu Leu Ala Asp Met
               1350                1355                1360

gcc acc acc atc atg gag act acc acc acc atg gcc act acc acc acg     4245
Ala Thr Thr Ile Met Glu Thr Thr Thr Thr Met Ala Thr Thr Thr Thr
           1365                1370                1375

cgc cgg ggc cgc tcc ccc aca ctg agg gac agc acc acc cag aac aca     4293
Arg Arg Gly Arg Ser Pro Thr Leu Arg Asp Ser Thr Thr Gln Asn Thr
       1380                1385                1390

gat gac ctg ctg gtg gcc tct gct gag tgt cca agc gat gat gag gac     4341
Asp Asp Leu Leu Val Ala Ser Ala Glu Cys Pro Ser Asp Asp Glu Asp
   1395                1400                1405

ctg gag gag tgt gag ccc agt act gga gga gag tta ata ttg ccc att     4389
Leu Glu Glu Cys Glu Pro Ser Thr Gly Gly Glu Leu Ile Leu Pro Ile
1410                1415                1420                1425

atc acg gag gac tcc tta gac ccc cct ccc gtg gcc acc cga tcc ccc     4437
Ile Thr Glu Asp Ser Leu Asp Pro Pro Pro Val Ala Thr Arg Ser Pro
               1430                1435                1440
```

```
ttc gtg ccc ccg ccc cct acc ttc tac ccc ttc ctc acg gga gtg ggc     4485
Phe Val Pro Pro Pro Pro Thr Phe Tyr Pro Phe Leu Thr Gly Val Gly
          1445                1450                1455

gcc acc caa gac acg ctg ccc ccg ccc gcc gcg cgc cgc ccg ccc tct     4533
Ala Thr Gln Asp Thr Leu Pro Pro Pro Ala Ala Arg Arg Pro Pro Ser
      1460                1465                1470

ggg ggc ccg tgc cag gcc gag cgg gac gac agc gac tgc gag gag ccc     4581
Gly Gly Pro Cys Gln Ala Glu Arg Asp Asp Ser Asp Cys Glu Glu Pro
   1475                1480                1485

atc gag gcc tcg ggc ttc gcc tcc ggg gag gtc ttt gac tcc agc ctc     4629
Ile Glu Ala Ser Gly Phe Ala Ser Gly Glu Val Phe Asp Ser Ser Leu
1490                1495                1500                1505

ccc ccc acg gac gac gag gac ttt tac acc acc ttt ccc ctg gtc acg     4677
Pro Pro Thr Asp Asp Glu Asp Phe Tyr Thr Thr Phe Pro Leu Val Thr
               1510                1515                1520

gac cgc acc acc ctc ctg tca ccc cgc aaa ccc gct ccc cgg ccc aac     4725
Asp Arg Thr Thr Leu Leu Ser Pro Arg Lys Pro Ala Pro Arg Pro Asn
          1525                1530                1535

ctc agg aca gat ggg gcc acg ggc gcc cct ggg gtg ctg ttt gcc ccc     4773
Leu Arg Thr Asp Gly Ala Thr Gly Ala Pro Gly Val Leu Phe Ala Pro
      1540                1545                1550

tcc gcc ccg gcc ccc aac ctg ccg gcg gga aaa atg aac cac cga gac     4821
Ser Ala Pro Ala Pro Asn Leu Pro Ala Gly Lys Met Asn His Arg Asp
   1555                1560                1565

ccg ctt cag ccc ttg ctg gag aac ccg ccc ttg ggg ccc ggg gcc ccc     4869
Pro Leu Gln Pro Leu Leu Glu Asn Pro Pro Leu Gly Pro Gly Ala Pro
1570                1575                1580                1585

acg tcc ttt gag ccg cgg agg ccc cct ccc ctg cgc ccc ggc gtg acc     4917
Thr Ser Phe Glu Pro Arg Arg Pro Pro Pro Leu Arg Pro Gly Val Thr
               1590                1595                1600

tca gcc ccc ggc ttc ccc cat ctg ccc aca gcc aac ccc aca ggg cct     4965
Ser Ala Pro Gly Phe Pro His Leu Pro Thr Ala Asn Pro Thr Gly Pro
          1605                1610                1615

ggg gag cgg ggc ccg ccg ggc gca gtg gag gtg atc cgg gag tcc agc     5013
Gly Glu Arg Gly Pro Pro Gly Ala Val Glu Val Ile Arg Glu Ser Ser
      1620                1625                1630

agc acc acg ggc atg gtg gtg ggc att gtg gcg gcg gcg gcg ctc tgc     5061
Ser Thr Thr Gly Met Val Val Gly Ile Val Ala Ala Ala Ala Leu Cys
   1635                1640                1645

atc ctc atc ctc ctc tac gcc atg tat aag tac cgc aat cgt gat gag     5109
Ile Leu Ile Leu Leu Tyr Ala Met Tyr Lys Tyr Arg Asn Arg Asp Glu
1650                1655                1660                1665

ggc tcc tac cag gtg gac cag agc cga aac tac atc agt aac tcg gcc     5157
```

```
Gly Ser Tyr Gln Val Asp Gln Ser Arg Asn Tyr Ile Ser Asn Ser Ala
                1670                1675                1680

cag agc aat ggg gcg gtg gtg aaa gag aag gcc ccg gct gcc ccc aag     5205
Gln Ser Asn Gly Ala Val Val Lys Glu Lys Ala Pro Ala Ala Pro Lys
            1685                1690                1695

acg ccc agc aag gcc aag aag aac aaa gac aag gag tat tat gtc tga     5253
Thr Pro Ser Lys Ala Lys Lys Asn Lys Asp Lys Glu Tyr Tyr Val  *
        1700                1705                1710

gcccccggca ctgcgcccca ctgccagctg cccctcctgg gagggcccgg gaggagggtg   5313

ccaccctctc cctgccaggg gcctggggac cctctccctg gctgcctcag gcttctctta   5373

cgaagaggaa acgcaaaaaa agaaaaggaa aaaccccgtg ctcgcccccc tcctcctgcc   5433

gtccactgcg cggcctcgtc agtcccgggg ctgactgtcc ctctcagctc tgcgcctgcc   5493

aggcagggca cgtgctcaca gccctgggtt gatttatttt tttaaggggg tagttttatt   5553

ttggtggggt tgggtgggaa ggaaggctgg gggttttgta aagtgtccac tgctcgtcct   5613

gttaattttc ctcaattttt cttcttcttc cttctgtccc tcctgccttc cttctcttcc   5673

caagccctcc aatccccatc ccaggcttgc tgtgtctcac tgtccccacc ctccttccct   5733

acttcttttt ttgtgtgtct ggtttctccc ttcctttcct ccctttgggt ttccagagtc   5793

ggtgggagaa gggcgggagg gtgggcccga gtggcccagt gggtgggtgg ggtggggtgg   5853

ggcaagtgcc ccaactcccc tcaccaggag aggcacctgc ttggtgccgc ccagggaagg   5913

ggctcaggcc tgacggaagg cctgttctgt gtgtgccgcc gggcgacgtg cattgatggg   5973

gaagctgctg gaggagcagg ggtgggggggt gggagggagg ggaaaggcaa atgcagatat   6033

atattacaga caaatactct agattccacg agcagcagcc tgtggcaccc gctgggcgcg   6093

ggcagcaggg aagagggagc aaggcattgt ccacagactg ctggggtcac ttctttgccc   6153

acgggctccc tgctccccca gtttttttttc tctctttgtt aacaaatgtg tctgagtctt   6213

ggaaaacacc ccaaccccgg aaatgtgtgg gaaaaagaaa acaaaaactt tccaaattcc   6273

aaaaaaaaaa aaaaa                                                    6288
```

<210> 24
<211> 1016
<212> DNA
<213> Homo sapiens

<220>
<221> CDS

<222> (254)..(724)

<400> 24

```
ttaatacgat gactataggg aatttggccc tcgagcaaga attcggcacg aggctggctc      60

tgactcttca cccatcttca cccaggctgg ccccttggt gaaactacaa ctcccagggg     120

tctgtgcgcg agaaggcagg cgggtttttc taccggaagt ccgctctagc tctgggccct    180

acaactgcac cctgagccgg agctgcccag tcgccgcggg accggggccg ctggggtctg    240

gacgggggtc gcc   atg atc cgc ttt atc ctc atc cag aac cgg gca ggc      289
                Met Ile Arg Phe Ile Leu Ile Gln Asn Arg Ala Gly
                  1               5                  10

aag acg cgc ctg gcc aag tgg tac atg cag ttt gat gat gat gag aaa      337
Lys Thr Arg Leu Ala Lys Trp Tyr Met Gln Phe Asp Asp Asp Glu Lys
         15                  20                  25

cag aag ctg atc gag gag gtg cat gcc gtg gtc acc gtc cga gac gcc      385
Gln Lys Leu Ile Glu Glu Val His Ala Val Val Thr Val Arg Asp Ala
     30                  35                  40

aaa cac acc aac ttt gtg gag gtc ctg gca agc tcc gtt gct gac agc      433
Lys His Thr Asn Phe Val Glu Val Leu Ala Ser Ser Val Ala Asp Ser
 45                  50                  55                  60

ctc tct gtt ctg cag ttc cgg aac ttt aag atc att tac cgc cgc tat      481
Leu Ser Val Leu Gln Phe Arg Asn Phe Lys Ile Ile Tyr Arg Arg Tyr
                 65                  70                  75

gct ggc ctc tac ttc tgc atc tgt gtg gat gtc aat gac aac aac ctg      529
Ala Gly Leu Tyr Phe Cys Ile Cys Val Asp Val Asn Asp Asn Asn Leu
             80                  85                  90

gct tac ctg gag gcc att cac aac ttc gtg gag gtc tta aac gaa tat      577
Ala Tyr Leu Glu Ala Ile His Asn Phe Val Glu Val Leu Asn Glu Tyr
         95                  100                 105

ttc cac aat gtc tgt gaa ctg gac ctg gtg ttc aac ttc tac aag gtt      625
Phe His Asn Val Cys Glu Leu Asp Leu Val Phe Asn Phe Tyr Lys Val
    110                 115                 120

tac acg gtc gtg gac gag atg ttc ctg gct ggc gaa atc cga gag acc      673
Tyr Thr Val Val Asp Glu Met Phe Leu Ala Gly Glu Ile Arg Glu Thr
125                 130                 135                 140

agc cag acg aag gtg ctg aaa cag ctg ctg atg cta cag tcc ctg gag      721
Ser Gln Thr Lys Val Leu Lys Gln Leu Leu Met Leu Gln Ser Leu Glu
                145                 150                 155

tga gggc aggcgagccc caccccggcc ccggcccctc ctggactcgc ctgctcgctt      778
 *


ccccttccca ggcccgtggc caacccagca gtccttccct cagctgccta ggaggaaggg    838
```

```
acccagctgg gtctgggcca caagggagga gactgcaccc cactgcctct gggccctggc    898

tgtgggcaga ggccaccgtg tgtgtcccga gtaaccgtgc cgttgtcgtg tgatgccata    958

agcgtctgtg cgtggagtcc ccaataaacc tgtggtcctg cctggcaaaa aaaaaaaa     1016
```

<210> 25
<211> 977
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (46)..(828)

<400> 25

```
agtgagcgct ggtggcggag ttaaagtcaa agcaggagag taatt    atg aat agc      54
                                                     Met Asn Ser
                                                       1

gca gcg gga ttc tca cac cta gac cgt cgc gag cgg gtt ctc aag tta      102
Ala Ala Gly Phe Ser His Leu Asp Arg Arg Glu Arg Val Leu Lys Leu
  5                  10                  15

ggg gag agt ttc gag aag cag ccg cgc tgc gcc ttc cac act gtg cgc      150
Gly Glu Ser Phe Glu Lys Gln Pro Arg Cys Ala Phe His Thr Val Arg
 20                  25                  30                  35

tat gac ttc aaa cct gct tct att gac act tct tct gaa gga tac ctt      198
Tyr Asp Phe Lys Pro Ala Ser Ile Asp Thr Ser Ser Glu Gly Tyr Leu
                 40                  45                  50

gag gtt ggt gaa ggt gaa cag gtg acc ata act ctg cca aat ata gaa      246
Glu Val Gly Glu Gly Glu Gln Val Thr Ile Thr Leu Pro Asn Ile Glu
             55                  60                  65

ggt tca act cca cca gta act gtt ttc aaa ggt tca aaa aaa cct tac      294
Gly Ser Thr Pro Pro Val Thr Val Phe Lys Gly Ser Lys Lys Pro Tyr
         70                  75                  80

tta aaa gaa tgc att ttg att att aac cat gat act gga gaa tgt cgg      342
Leu Lys Glu Cys Ile Leu Ile Ile Asn His Asp Thr Gly Glu Cys Arg
  85                  90                  95

cta gaa aaa ctc agc agc aac atc act gta aaa aaa aca aga gtt gaa      390
Leu Glu Lys Leu Ser Ser Asn Ile Thr Val Lys Lys Thr Arg Val Glu
100                 105                 110                 115

gga agc agt aaa att cag tat cgt aaa gaa caa cag caa caa caa atg      438
Gly Ser Ser Lys Ile Gln Tyr Arg Lys Glu Gln Gln Gln Gln Gln Met
                120                 125                 130

tgg aat tca gcc agg act ccc aat ctt gta aaa cat tct cca tct gaa      486
Trp Asn Ser Ala Arg Thr Pro Asn Leu Val Lys His Ser Pro Ser Glu
```

```
            135                 140                 145

gat aag atg tcc cca gca tct cca ata gat gat atc gaa aga gaa ctg      534
Asp Lys Met Ser Pro Ala Ser Pro Ile Asp Asp Ile Glu Arg Glu Leu
        150                 155                 160

aag gca gaa gct agt cta atg gac cag atg agt agt tgt gat agt tca      582
Lys Ala Glu Ala Ser Leu Met Asp Gln Met Ser Ser Cys Asp Ser Ser
    165                 170                 175

tca gat tcc aaa agt tca tca tct tca agt agt gag gat agt tct agt      630
Ser Asp Ser Lys Ser Ser Ser Ser Ser Ser Ser Glu Asp Ser Ser Ser
180                 185                 190                 195

gac tca gaa gat gaa gat tgc aaa tcc tct act tct gat aca ggg aat      678
Asp Ser Glu Asp Glu Asp Cys Lys Ser Ser Thr Ser Asp Thr Gly Asn
                200                 205                 210

tgt gtc tca gga cat cct acc atg aca cag tac agg att cct gat ata      726
Cys Val Ser Gly His Pro Thr Met Thr Gln Tyr Arg Ile Pro Asp Ile
            215                 220                 225

gat gcc agt cat aat aga ttt cga gac aac agt ggc ctt ctg atg aat      774
Asp Ala Ser His Asn Arg Phe Arg Asp Asn Ser Gly Leu Leu Met Asn
        230                 235                 240

act tta aga aat gat ttg cag ctg agt gaa tca gga agt gac agt gat      822
Thr Leu Arg Asn Asp Leu Gln Leu Ser Glu Ser Gly Ser Asp Ser Asp
    245                 250                 255

gac tga agaaatattt agctataaat aaaaatttat acagcatgta taatttattt      878
Asp  *
260

tgtattaaca ataaaaattc ctaagactga gggaaatatg tcttaacttt tgatgataaa    938

agaaattaaa tttgattcag aaatttcaaa aaaaaaaaa                           977
```


<210> 26
<211> 3185
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (37)..(1257)

<400> 26

```
accccggcgg ccctgcccgc ccctccctcc agcatc  atg gcc agc cca aga acc      54
                                         Met Ala Ser Pro Arg Thr
                                           1               5

agg aag gtt ctt aaa gaa gtc agg gtg cag gat gag aac aac gtt tgt      102
Arg Lys Val Leu Lys Glu Val Arg Val Gln Asp Glu Asn Asn Val Cys
```

```
                10                  15                  20

ttt gag tgt ggc gcg ttc aat cct cag tgg gtc agt gtg acc tac ggc      150
Phe Glu Cys Gly Ala Phe Asn Pro Gln Trp Val Ser Val Thr Tyr Gly
         25                  30                  35

atc tgg atc tgc ctg gag tgc tcg ggg aga cac cgc ggg ctt ggg gtt      198
Ile Trp Ile Cys Leu Glu Cys Ser Gly Arg His Arg Gly Leu Gly Val
     40                  45                  50

cac ctc agc ttt gtg cgc tct gtt act atg gac aag tgg aag gac att      246
His Leu Ser Phe Val Arg Ser Val Thr Met Asp Lys Trp Lys Asp Ile
 55                  60                  65                  70

gag ctt gag aag atg aaa gct ggt ggg aat gct aag ttc cga gag ttc      294
Glu Leu Glu Lys Met Lys Ala Gly Gly Asn Ala Lys Phe Arg Glu Phe
                 75                  80                  85

ctg gag tct cag gag gat tac gat cct tgc tgg tcc ttg cag gag aag      342
Leu Glu Ser Gln Glu Asp Tyr Asp Pro Cys Trp Ser Leu Gln Glu Lys
             90                  95                 100

tac aac agc aga gcc gcg gcc ctc ttt agg gat aag gtg gtc gct ctg      390
Tyr Asn Ser Arg Ala Ala Ala Leu Phe Arg Asp Lys Val Val Ala Leu
        105                 110                 115

gcc gaa ggc aga gag tgg tct ctg gag tca tca cct gcc cag aac tgg      438
Ala Glu Gly Arg Glu Trp Ser Leu Glu Ser Ser Pro Ala Gln Asn Trp
    120                 125                 130

acc cca cct cag ccc agg acg ctg ccg tcc atg gtg cac cga gtc tct      486
Thr Pro Pro Gln Pro Arg Thr Leu Pro Ser Met Val His Arg Val Ser
135                 140                 145                 150

ggc cag ccg cag agt gtg acc gcc tcc tcg gac aag gct ttt gaa gac      534
Gly Gln Pro Gln Ser Val Thr Ala Ser Ser Asp Lys Ala Phe Glu Asp
                155                 160                 165

tgg ctg aat gat gac ctc ggc tcc tat caa ggg gcc cag ggg aat cgc      582
Trp Leu Asn Asp Asp Leu Gly Ser Tyr Gln Gly Ala Gln Gly Asn Arg
            170                 175                 180

tac gtg ggg ttt ggg aac acg cca ccg cct cag aag aaa gaa gat gac      630
Tyr Val Gly Phe Gly Asn Thr Pro Pro Pro Gln Lys Lys Glu Asp Asp
        185                 190                 195

ttc ctc aac aac gcc atg tcc tcc ctg tac tcg ggc tgg agc agc ttc      678
Phe Leu Asn Asn Ala Met Ser Ser Leu Tyr Ser Gly Trp Ser Ser Phe
    200                 205                 210

acc act gga gcc agc cgg ttt gcc tcg gca gcc aag gag ggc gct aca      726
Thr Thr Gly Ala Ser Arg Phe Ala Ser Ala Ala Lys Glu Gly Ala Thr
215                 220                 225                 230

aag ttt gga tcc caa gcg agt cag aag gcg tcc gag ctg ggc cac agc      774
Lys Phe Gly Ser Gln Ala Ser Gln Lys Ala Ser Glu Leu Gly His Ser
                235                 240                 245
```

```
ctg aac gag aac gtc ctc aag cct gcg cag gag aag gtg aag gag gga      822
Leu Asn Glu Asn Val Leu Lys Pro Ala Gln Glu Lys Val Lys Glu Gly
            250                 255                 260

aag att ttt gat gat gtc tcc agt ggg gtc tct cag ttg gcg tcc aag      870
Lys Ile Phe Asp Asp Val Ser Ser Gly Val Ser Gln Leu Ala Ser Lys
        265                 270                 275

gtc cag gga gtc ggt agt aag gga tgg cgg gac gtc acc acc ttt ttt      918
Val Gln Gly Val Gly Ser Lys Gly Trp Arg Asp Val Thr Thr Phe Phe
    280                 285                 290

tcg ggg aaa gca gag ggc ccc ttg gac agc ccc tcg gag ggc cac agt      966
Ser Gly Lys Ala Glu Gly Pro Leu Asp Ser Pro Ser Glu Gly His Ser
295                 300                 305                 310

tat cag aac agc ggt ctg gac cac ttc caa aac agc aac ata gac cag     1014
Tyr Gln Asn Ser Gly Leu Asp His Phe Gln Asn Ser Asn Ile Asp Gln
                315                 320                 325

agc ttc tgg gag acc ttt gga agt gct gag ccc acc aag acc cgc aag     1062
Ser Phe Trp Glu Thr Phe Gly Ser Ala Glu Pro Thr Lys Thr Arg Lys
            330                 335                 340

tcc ccg agc agc gac agc tgg acg tgc gcg gac acc tcc acc gag agg     1110
Ser Pro Ser Ser Asp Ser Trp Thr Cys Ala Asp Thr Ser Thr Glu Arg
        345                 350                 355

agg agc tcg gac agc tgg gag gtg tgg ggc tcg gcc tcc acc aac agg     1158
Arg Ser Ser Asp Ser Trp Glu Val Trp Gly Ser Ala Ser Thr Asn Arg
    360                 365                 370

aac agc aac agc gac ggc ggg gag ggc ggg gag ggc acc aag aag gca     1206
Asn Ser Asn Ser Asp Gly Gly Glu Gly Gly Glu Gly Thr Lys Lys Ala
375                 380                 385                 390

gtg ccg ccg gcc gtg ccc act gat gat ggc tgg gac aac cag aac tgg     1254
Val Pro Pro Ala Val Pro Thr Asp Asp Gly Trp Asp Asn Gln Asn Trp
                395                 400                 405

tag ggcc cactgcgccc ccgtccccag cgcccccggg cgacttcgtg tttgcactct     1311
 *

gccctcgtcg ttcctcctcc ttccatttga cccaagaatc agcaactgca gtgtgaggac   1371

agcgtctcgg gaggcaggac cctagggaga cccgggtgtg cgccgcctgc gcgtggggag   1431

tcttcggtgc gtgggggcgg cttgctgtcc agcctgtgtg ggggccgtcc cgtcccacac   1491

tcccctgggc attcttggac tcaaggccgg ggctctgcgt ggcttgctgg gaggtgggct   1551

gcagcacaga ggcctgtgac tgcgttccag cggccagttc actacgcagt atctctgggg   1611

cctgggacca gccacgtgcc gagctgtcag cgacgtgagg tgtcccttct cgttgagata   1671
```

tttaactttg gttttgctct agttctttct ttttgaagag agtgactgga gtggtaaaga 1731

tggaaatgct ggaaatgata ctggcgctca cgctgccatc cgaccaccct cggctcccga 1791

gtccacgcct gcctgggcct gtgctgtcag acccgcgtcg gtcgtaaccc tctgtggctc 1851

ccctgcatca gcaccgtccc accaccaagt tcaccaggtt caccagacac ggcctccaca 1911

atagccacac ccacacctga gctgttctca gtgctggaac ttgaccatcc tggaacaccc 1971

tggaagaaaa aggagcgcag ggtgggccct cggccctgat gcaggagggt gcgatagcgg 2031

acgtggccag gcaggagggg ccgggttcag gagctgagca ggggatgcct gtgcgtggtg 2091

cctgggtcta gggaagctcc agccccagga tggggctgcc ctgcacaccg gtgcccgcca 2151

catgccaacc ctcacctccc cgaggactgg atgatgtgct gccacgtgtg actcgtctcc 2211

cttgtctgcc ctgtgtgacc ctcagtcttg gccagccatg catgcgcccg aagctcgtgc 2271

agtttgtacg tgaggtgctc tcctccctgc caccatgctc atcactctgg ccttggccat 2331

gctccctggt caccccactt cccggtcgcc gtctgcagca ctcctggagc agcctgggcc 2391

cttcagcccc tgtgctcgtc ccaccctagg gactcagcca cttgcagaac aggatgggac 2451

cgagatttca gcgagccctc ctggcgcccg gtcctccctg tgggcaccag ccctcttggt 2511

agctggtgtg gagggccggt gtccttggct gccacggagg gatttgatca ccgaagcagc 2571

cacctgctgt agttggacct gaggtcagag gcggggcatc agaggctcaa ggtgctgaga 2631

agccaccggg aaagcagcca gcacaaaggg cccaggaagc cagcccccga gagctgagcg 2691

tgggggtctt tgagtgtctt tctccaagct gagacgtggg cggccgcgtg gtatctcccg 2751

agggctgctt ggaccctggt gggctgagtg ctccgaggag gggtggactc caccttggac 2811

agtgggatgt ggtgttccac atgtgcctgt ttccacgcca gcaccttgac ttggcagcat 2871

ggagccaagg tctgtccccg cccaggaggg tgccttcctc gggggtaggg ggacggccca 2931

ctctgcccca gggagtccct tttgatggga agtgcagtca gcagcgtgga ggtgtctggg 2991

ccaccttcag aaggtggatg tggtggccga gaccccgtcc acggagggtg atggcctttc 3051

ccttctgcag gtgcgggcag gtgggcctgg gaccggtgct ggggcctctc cttgctgtgt 3111

gtgagggccc aggtggaagg cgcggacctg acagcattcc aataaagcat acgggaacat 3171

gcaaaaaaaa aaaa 3185

<210> 27
<211> 1850

<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (596)..(1840)

<400> 27
```
ttttttgtatg caccacgggc ggcggtggtc ggtgcgggag gagggagggg agcttgcggg     60

cccgagaggg ggcgacggcg gcggcggtgg cctgaggagg cccgagcggc ggcggtggcg    120

gcgaaggccg aggcgtctag gtgtttttgg aagagctgca gccctcttct cacagatgag    180

ctacgaggag atgatgacac tgactgagca gcacctggag tctcagaacg tcaccaaagg    240

tgcccgccac aagatagccc tgagcatcca gaagctgcgt gagagacaga gcgtcctcaa    300

gtccctagag aaggatgtgc tggaaggcgg gaacctacga aacgctctgc aggagctgca    360

gcagatcatc atcactccca tcaaggccta cagtgtcctc caggccaccg tggctgccgc    420

caccaccacc cctactgcca aggatggggc cccgggggaa ccaccgctgc caggtgctga    480

gcctccccta gcccaccccg gcacagacaa aggcaccgag gccagagccg ggaccatgtg    540

acggcgctgg ccctcgccac cgccgtcccc cgaccctggc cccaggcccg gcacc atg      598
                                                              Met
                                                                1

atg ttc cga gac cag gtg ggc atc ctc gct ggc tgg ttc aaa ggc tgg      646
Met Phe Arg Asp Gln Val Gly Ile Leu Ala Gly Trp Phe Lys Gly Trp
                5                  10                  15

aat gag tgt gag cag aca gtg gcc ctc ctg tca ctt ccg aaa cgg gtc      694
Asn Glu Cys Glu Gln Thr Val Ala Leu Leu Ser Leu Pro Lys Arg Val
         20                  25                  30

acc cgt acc cag gcc cgc ttc ctg cag ctc tgc ctg gag cac tca ctg      742
Thr Arg Thr Gln Ala Arg Phe Leu Gln Leu Cys Leu Glu His Ser Leu
     35                  40                  45

gcg gac tgc aat gac atc cac ctg ctg gag tcg gag gcc aac agt gct      790
Ala Asp Cys Asn Asp Ile His Leu Leu Glu Ser Glu Ala Asn Ser Ala
 50                  55                  60                  65

gcc atc gtc agc cag tgg cag cag gag tcc aaa gag aag gtg gtg tcc      838
Ala Ile Val Ser Gln Trp Gln Gln Glu Ser Lys Glu Lys Val Val Ser
                70                  75                  80

ctc ctg ctg tcc cac ctt ccc ctg ctt cag cca ggc aac aca gag gcc      886
Leu Leu Leu Ser His Leu Pro Leu Leu Gln Pro Gly Asn Thr Glu Ala
            85                  90                  95

aag tcg gag tac atg agg cta ctg cag aaa gtg ctg gcc tac tca atc      934
Lys Ser Glu Tyr Met Arg Leu Leu Gln Lys Val Leu Ala Tyr Ser Ile
        100                 105                 110
```

```
gag agc aat gct ttc atc gag gag agt cgc cag ctg ctt tcc tat gcc      982
Glu Ser Asn Ala Phe Ile Glu Glu Ser Arg Gln Leu Leu Ser Tyr Ala
    115                 120                 125

ctc atc cac cca gcc acc aca ctg gag gac cgc aac gca ctg gcc ctc     1030
Leu Ile His Pro Ala Thr Thr Leu Glu Asp Arg Asn Ala Leu Ala Leu
130                 135                 140                 145

tgg ctg agc cac ctg gaa gag cgg ttg gct agt ggc ttc cgc tcc cgg     1078
Trp Leu Ser His Leu Glu Glu Arg Leu Ala Ser Gly Phe Arg Ser Arg
                150                 155                 160

cca gag ccc tcc tac cat tca cgt caa ggc tca gat gag tgg ggg ggc     1126
Pro Glu Pro Ser Tyr His Ser Arg Gln Gly Ser Asp Glu Trp Gly Gly
            165                 170                 175

cct gca gag cta ggc cct ggg gag gca ggg cca ggc tgg cag gac aag     1174
Pro Ala Glu Leu Gly Pro Gly Glu Ala Gly Pro Gly Trp Gln Asp Lys
        180                 185                 190

cca ccc cgg gaa aat gga cac gtg ccc ttc cac cca tcc agc tca gtg     1222
Pro Pro Arg Glu Asn Gly His Val Pro Phe His Pro Ser Ser Ser Val
    195                 200                 205

ccg cca gcc atc aac agt att ggg agc aat gca aac aca ggt ctc ccc     1270
Pro Pro Ala Ile Asn Ser Ile Gly Ser Asn Ala Asn Thr Gly Leu Pro
210                 215                 220                 225

tgc caa atc cac cct agc cca ctg aag cgc tcc atg tca ctc atc cct     1318
Cys Gln Ile His Pro Ser Pro Leu Lys Arg Ser Met Ser Leu Ile Pro
                230                 235                 240

aca agc ccc cag gtc cct ggt gag tgg ccg agt cca gag gag ctt ggg     1366
Thr Ser Pro Gln Val Pro Gly Glu Trp Pro Ser Pro Glu Glu Leu Gly
            245                 250                 255

gcc cgg gct gct ttt acc acg ccc gat cac gca cct ctc tcg ccc cag     1414
Ala Arg Ala Ala Phe Thr Thr Pro Asp His Ala Pro Leu Ser Pro Gln
        260                 265                 270

agc agc gtg gcc tcc tct ggc agt gag cag aca gag gag cag ggc tcc     1462
Ser Ser Val Ala Ser Ser Gly Ser Glu Gln Thr Glu Glu Gln Gly Ser
    275                 280                 285

agc cgg aac acc ttc cag gag gat ggc agt ggc atg aaa gat gtg ccc     1510
Ser Arg Asn Thr Phe Gln Glu Asp Gly Ser Gly Met Lys Asp Val Pro
290                 295                 300                 305

tca tgg ctc aag agc ctc cgt ttg cac aag tat gca gcc ctc ttc tca     1558
Ser Trp Leu Lys Ser Leu Arg Leu His Lys Tyr Ala Ala Leu Phe Ser
                310                 315                 320

cag atg agc tac gag gag atg atg aca ctg act gag cag cac ctg gag     1606
Gln Met Ser Tyr Glu Glu Met Met Thr Leu Thr Glu Gln His Leu Glu
            325                 330                 335
```

```
tct cag aac gtc acc aaa ggt gcc cgc cac aag ata gcc ctg agc atc     1654
Ser Gln Asn Val Thr Lys Gly Ala Arg His Lys Ile Ala Leu Ser Ile
        340                 345                 350

cag aag ctg cgt gag aga cag agc gtc ctc aag tcc cta gag aag gat     1702
Gln Lys Leu Arg Glu Arg Gln Ser Val Leu Lys Ser Leu Glu Lys Asp
    355                 360                 365

gtg ctg gaa ggc ggg aac cta cga aac gct ctg cag gag ctg cag cag     1750
Val Leu Glu Gly Gly Asn Leu Arg Asn Ala Leu Gln Glu Leu Gln Gln
370                 375                 380                 385

atc atc atc act ccc atc aag gcc tac agt gtc ctc cag gcc acc gtg     1798
Ile Ile Ile Thr Pro Ile Lys Ala Tyr Ser Val Leu Gln Ala Thr Val
                390                 395                 400

gct gcc gcc acc ctg tat tgc ggc cgc tct aga gga tcc tag cttcgta     1847
Ala Ala Ala Thr Leu Tyr Cys Gly Arg Ser Arg Gly Ser  *
            405                 410                 415

cgc                                                                 1850
```

<210> 28
<211> 4781
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (813)..(3635)

<400> 28

```
tttgctgtgg aattcccggg tcgacccacg cgtccgcgga cgcgtgggtt tagaggtaag     60

tttgcctact ttgtcgtcta gtgggtaaaa ttttgcggag agcgttggat ctgggaagcg    120

ggatagggat ggatgggttc atttgagagc cacggcttaa agcggttgcg atcaggatgg    180

gacacaggtt tgtttgggga caacaaagat ggcatttgtg agtgttttga agcaacccgt    240

actgattaca tctttctccc ttgtgttcct tttatcccag gtttgaattt tctcggagaa    300

agacaggccg gccacgagga aaacagaaac aagccgcagc aacatctaag cccttgaaag    360

gatcctgaga gaggggggaa agggaaaaca gcagccacca gcccaaccac ttgtgtcttc    420

tgccccttcc cacctatctt gcccacccca ccagcccacg ctgcttggga cttgaaatct    480

gtggccgaag gaccgtcact acataacttc aaaaataatc aaccaccctc ccttcccaaa    540

ccacccaaat tcactcatcc agcgtttact tttttgaatc cactcagaac ttttttctgc    600

gacccccctc cctaaatgga gttgggtggg ggggaaatga atactgagtt ggcctttatt    660
```

```
ttttaaaaga ctttttgatc caatgaggcc ccctaaataa ttgagttttg ggtcctggtt    720

ggttgtttta ttttttttcc tccaaaattt tacccccctcc cccctgagcc cgaggtgctg    780

acgtcgcaaa aaaattggat aaaaccacca tc  atg ggt tcg ggt ccc ata gac      833
                                        Met Gly Ser Gly Pro Ile Asp
                                          1               5

ccc aaa gaa ctt ctc aag ggc ctg gac agc ttc ctt aac cga gat ggg      881
Pro Lys Glu Leu Leu Lys Gly Leu Asp Ser Phe Leu Asn Arg Asp Gly
         10                  15                  20

gaa gtc aaa agt gtg gat ggg att tcc aag atc ttc agt ttg atg aag      929
Glu Val Lys Ser Val Asp Gly Ile Ser Lys Ile Phe Ser Leu Met Lys
     25                  30                  35

gaa gca cga aag atg gtg agt cga tgc act tac ttg aac att ctc ctg      977
Glu Ala Arg Lys Met Val Ser Arg Cys Thr Tyr Leu Asn Ile Leu Leu
 40                  45                  50                  55

cag acc cgt tca cca gaa ata ttg gtc aaa ttt att gac gtt ggc ggc     1025
Gln Thr Arg Ser Pro Glu Ile Leu Val Lys Phe Ile Asp Val Gly Gly
                 60                  65                  70

tac aaa ctt ctt aac aat tgg ctg acg tat tca aag aca acc aac aac     1073
Tyr Lys Leu Leu Asn Asn Trp Leu Thr Tyr Ser Lys Thr Thr Asn Asn
             75                  80                  85

att ccc ctc ctc cag caa att cta ctg acc ctg cag cat cta ccg ctc     1121
Ile Pro Leu Leu Gln Gln Ile Leu Leu Thr Leu Gln His Leu Pro Leu
         90                  95                  100

act gta gac cat ctc aag cag aac aac aca gct aaa ctg gtg aag cag     1169
Thr Val Asp His Leu Lys Gln Asn Asn Thr Ala Lys Leu Val Lys Gln
    105                 110                 115

ctg agc aag tca agt gag gat gaa gag ctc cgg aaa ttg gcc tca gtc     1217
Leu Ser Lys Ser Ser Glu Asp Glu Glu Leu Arg Lys Leu Ala Ser Val
120                 125                 130                 135

ctt gtc agc gac tgg atg gct gtc atc cgc tct cag agc agt acc cag     1265
Leu Val Ser Asp Trp Met Ala Val Ile Arg Ser Gln Ser Ser Thr Gln
                140                 145                 150

cct gct gag aaa gat aag aag aaa cgt aaa gat gaa gga aaa agt cga     1313
Pro Ala Glu Lys Asp Lys Lys Lys Arg Lys Asp Glu Gly Lys Ser Arg
            155                 160                 165

acc acc ctt cct gag cga cct ttg aca gag gtg aag gct gag acc cgg     1361
Thr Thr Leu Pro Glu Arg Pro Leu Thr Glu Val Lys Ala Glu Thr Arg
        170                 175                 180

gct gag gag gcc cca gag aag aag agg gag aag ccc aag tct ctt cgc     1409
Ala Glu Glu Ala Pro Glu Lys Lys Arg Glu Lys Pro Lys Ser Leu Arg
    185                 190                 195

acc aca gca ccc agt cat gcc aag ttc cgt tcc act gga cta gag ctg     1457
```

```
Thr Thr Ala Pro Ser His Ala Lys Phe Arg Ser Thr Gly Leu Glu Leu
200                 205                 210                 215

gag aca cca tcc ttg gtg cct gtg aag aag aat gcc agc aca gtg gtg     1505
Glu Thr Pro Ser Leu Val Pro Val Lys Lys Asn Ala Ser Thr Val Val
                220                 225                 230

gtt tct gac aag tac aac ctt aaa ccc atc ccc ctc aaa cgt cag agc     1553
Val Ser Asp Lys Tyr Asn Leu Lys Pro Ile Pro Leu Lys Arg Gln Ser
            235                 240                 245

aac gta gct gct cca gga gat gcc act ccc cct gca gag aag aaa tac     1601
Asn Val Ala Ala Pro Gly Asp Ala Thr Pro Pro Ala Glu Lys Lys Tyr
        250                 255                 260

aag cca ctc aac aca aca cct aat gcc acc aaa gag atc aaa gtg aag     1649
Lys Pro Leu Asn Thr Thr Pro Asn Ala Thr Lys Glu Ile Lys Val Lys
    265                 270                 275

atc atc ccg cca cag cct atg gag ggc ctg ggc ttt ctg gat gct ctt     1697
Ile Ile Pro Pro Gln Pro Met Glu Gly Leu Gly Phe Leu Asp Ala Leu
280                 285                 290                 295

aat tca gcc cct gtt cca ggc atc aaa att aag aag aaa aaa aaa gta     1745
Asn Ser Ala Pro Val Pro Gly Ile Lys Ile Lys Lys Lys Lys Lys Val
                300                 305                 310

ctg tca cct acg gct gcc aag cca agc ccc ttt gaa ggg aaa acg agc     1793
Leu Ser Pro Thr Ala Ala Lys Pro Ser Pro Phe Glu Gly Lys Thr Ser
            315                 320                 325

aca gaa cca agc aca gcc aaa cct tct tcc cca gaa cca gca cca cct     1841
Thr Glu Pro Ser Thr Ala Lys Pro Ser Ser Pro Glu Pro Ala Pro Pro
        330                 335                 340

tct gag gca atg gac gca gac cgt cca ggc acc ccg gtt ccc cct gtt     1889
Ser Glu Ala Met Asp Ala Asp Arg Pro Gly Thr Pro Val Pro Pro Val
    345                 350                 355

gaa gtc ccg gag ctc atg gat aca gcc tct ttg gag cca gga gct ctg     1937
Glu Val Pro Glu Leu Met Asp Thr Ala Ser Leu Glu Pro Gly Ala Leu
360                 365                 370                 375

gat gcc aag cca gtg gag agt cct gga gat cct aac caa ctg acc cgg     1985
Asp Ala Lys Pro Val Glu Ser Pro Gly Asp Pro Asn Gln Leu Thr Arg
                380                 385                 390

aaa ggc agg aag agg aaa agt gtg aca tgg cct gag gaa ggc aaa ctg     2033
Lys Gly Arg Lys Arg Lys Ser Val Thr Trp Pro Glu Glu Gly Lys Leu
            395                 400                 405

aga gaa tat ttc tat ttt gaa ttg gat gaa act gaa cga gta aat gtg     2081
Arg Glu Tyr Phe Tyr Phe Glu Leu Asp Glu Thr Glu Arg Val Asn Val
        410                 415                 420

aat aag atc aag gac ttt ggt gag gcg gct aag cga gag ata ctg tca     2129
Asn Lys Ile Lys Asp Phe Gly Glu Ala Ala Lys Arg Glu Ile Leu Ser
```

```
    425                 430                 435

gac cga cat gca ttt gag aca gcg cgg cgt ctg agc cat gat aac atg     2177
Asp Arg His Ala Phe Glu Thr Ala Arg Arg Leu Ser His Asp Asn Met
440                 445                 450                 455

gag gag aag gtg ccc tgg gtg tgc ccc cgg ccc ctg gtt ctg ccc tca     2225
Glu Glu Lys Val Pro Trp Val Cys Pro Arg Pro Leu Val Leu Pro Ser
                460                 465                 470

cct ctt gtc acc cct gga agc aat agt cag gag cga tat atc cag gct     2273
Pro Leu Val Thr Pro Gly Ser Asn Ser Gln Glu Arg Tyr Ile Gln Ala
            475                 480                 485

gag cgg gag aag gga atc ctt cag gag ctc ttc ctg aac aag gag agt     2321
Glu Arg Glu Lys Gly Ile Leu Gln Glu Leu Phe Leu Asn Lys Glu Ser
        490                 495                 500

cct cat gag cct gat cct gag ccc tac gag ccc ata ccc cct aaa ctc     2369
Pro His Glu Pro Asp Pro Glu Pro Tyr Glu Pro Ile Pro Pro Lys Leu
    505                 510                 515

atc ccc cta gat gag gag tgt tcc atg gat gag act ccg tat gtt gag     2417
Ile Pro Leu Asp Glu Glu Cys Ser Met Asp Glu Thr Pro Tyr Val Glu
520                 525                 530                 535

act ctg gaa cct ggg ggg tca ggt ggc tca cct gat ggg gca gga ggc     2465
Thr Leu Glu Pro Gly Gly Ser Gly Gly Ser Pro Asp Gly Ala Gly Gly
                540                 545                 550

tcc aag ttg cct cca gtt ctg gcc aat ctt atg gga agc atg ggt gct     2513
Ser Lys Leu Pro Pro Val Leu Ala Asn Leu Met Gly Ser Met Gly Ala
            555                 560                 565

gga aag ggc ccc caa ggc cct gga gga gga ggc att aat gtc caa gag     2561
Gly Lys Gly Pro Gln Gly Pro Gly Gly Gly Gly Ile Asn Val Gln Glu
        570                 575                 580

atc ctc acc tcc atc atg ggt agc cca aac agt cat cct tca gag gaa     2609
Ile Leu Thr Ser Ile Met Gly Ser Pro Asn Ser His Pro Ser Glu Glu
    585                 590                 595

cta ctg aaa caa cca gac tat tcg gac aag atc aag cag atg ctg gtg     2657
Leu Leu Lys Gln Pro Asp Tyr Ser Asp Lys Ile Lys Gln Met Leu Val
600                 605                 610                 615

cca cat gga ctc cta ggc cct ggc cca ata gcc aat ggt ttc cca cca     2705
Pro His Gly Leu Leu Gly Pro Gly Pro Ile Ala Asn Gly Phe Pro Pro
                620                 625                 630

ggg ggt cct ggg ggc ccc aag ggc atg cag cac ttt ccc cct gga cct     2753
Gly Gly Pro Gly Gly Pro Lys Gly Met Gln His Phe Pro Pro Gly Pro
            635                 640                 645

ggg gga cct atg cca ggt ccc cat gga ggc cct ggt ggg cca gtg ggt     2801
Gly Gly Pro Met Pro Gly Pro His Gly Gly Pro Gly Gly Pro Val Gly
        650                 655                 660
```

```
cca cgt ctt ctg ggt cct cca ccc cct ccc cgg gga ggt gat ccc ttc     2849
Pro Arg Leu Leu Gly Pro Pro Pro Pro Arg Gly Gly Asp Pro Phe
    665                 670                 675

tgg gat ggc ccg ggc gac cct atg cgg ggt ggc cca atg cgg ggg ggt     2897
Trp Asp Gly Pro Gly Asp Pro Met Arg Gly Gly Pro Met Arg Gly Gly
680                 685                 690                 695

cca gga cca ggt cct gga cca tac cat aga ggc cga ggt ggc cga gga     2945
Pro Gly Pro Gly Pro Gly Pro Tyr His Arg Gly Arg Gly Gly Arg Gly
                700                 705                 710

gga aac gaa cct cct cct cct cct cct cca ttc cga ggc gcc aga gga     2993
Gly Asn Glu Pro Pro Pro Pro Pro Pro Phe Arg Gly Ala Arg Gly
            715                 720                 725

ggt cgc tct gga gga gga ccc cca aat gga cga ggg ggc cct ggt ggg     3041
Gly Arg Ser Gly Gly Gly Pro Pro Asn Gly Arg Gly Gly Pro Gly Gly
        730                 735                 740

ggc atg gtt gga ggt ggt ggg cat cgt cct cac gaa ggc cct ggt ggg     3089
Gly Met Val Gly Gly Gly Gly His Arg Pro His Glu Gly Pro Gly Gly
    745                 750                 755

ggc atg ggc aac agc agt gga cat cgt ccc cac gaa ggc cct ggc ggt     3137
Gly Met Gly Asn Ser Ser Gly His Arg Pro His Glu Gly Pro Gly Gly
760                 765                 770                 775

ggc atg gga agt ggg cat cgc ccc cat gaa ggc cct ggt ggt agc atg     3185
Gly Met Gly Ser Gly His Arg Pro His Glu Gly Pro Gly Gly Ser Met
                780                 785                 790

ggt ggg ggt gga gga cat cgt ccc cac gaa ggc cct ggc ggt ggc atc     3233
Gly Gly Gly Gly Gly His Arg Pro His Glu Gly Pro Gly Gly Gly Ile
            795                 800                 805

agt ggt ggc agt ggc cat cgt ccc cat gaa ggc cct ggc gga gga atg     3281
Ser Gly Gly Ser Gly His Arg Pro His Glu Gly Pro Gly Gly Gly Met
        810                 815                 820

ggt gcc ggt ggt gga cat cgc ccc cac gaa ggc cct ggc gga agc atg     3329
Gly Ala Gly Gly Gly His Arg Pro His Glu Gly Pro Gly Gly Ser Met
    825                 830                 835

ggt gga agt ggt gga cat cgt ccc cat gaa ggc cct gga cac ggg ggg     3377
Gly Gly Ser Gly Gly His Arg Pro His Glu Gly Pro Gly His Gly Gly
840                 845                 850                 855

ccc cat ggc cac cgg cct cat gat gtc cct ggt cac cga ggc cat gac     3425
Pro His Gly His Arg Pro His Asp Val Pro Gly His Arg Gly His Asp
                860                 865                 870

cat cga ggg ccg cca cct cat gag cac cgt ggc cat gat ggt cct ggc     3473
His Arg Gly Pro Pro Pro His Glu His Arg Gly His Asp Gly Pro Gly
            875                 880                 885
```

```
cac ggg gga ggg ggc cac cga ggg cac gat gga ggc cac agc cat gga     3521
His Gly Gly Gly Gly His Arg Gly His Asp Gly Gly His Ser His Gly
        890                 895                 900

gga gac atg tca aac cgc cct gtc tgc cga cat ttc atg atg aag ggc     3569
Gly Asp Met Ser Asn Arg Pro Val Cys Arg His Phe Met Met Lys Gly
    905                 910                 915

aac tgc cgc tat gag aac aac tgt gcc ttc tac cac ccg ggt gtc aat     3617
Asn Cys Arg Tyr Glu Asn Asn Cys Ala Phe Tyr His Pro Gly Val Asn
920                 925                 930                 935

ggg ccc ccc ctg ccc tag ggacca tttgcctgcc ctgttcacac aacccctgtg     3671
Gly Pro Pro Leu Pro  *
                940

gactgcagcc tcgctctttc caccctgtta tggcttctgt gaggcccatt ttccctttc   3731

cccagctgat gaggagccgg ccccctcagt tcccacttgc ttgggttcct gggggttttc   3791

tgatcactgg tgcgcattga tgtacatatt ttcctccagt ctggggagga gagagactgg   3851

aaacgttcct ggactgctga agaggagacc cagttggctt cacttttga gaagattcgc   3911

cctgtacccc aaacccctttt ccagtattac ccttaatgct tgagaaccta aagctggtta   3971

tcctggcgaa caccctacc cttctattgc gggtccccac atgcacacag aactctgaca   4031

caggatcagc tgcacttaag aaatcatccc agctaagttc attattcctc atggggtggg   4091

gagatgctga aaggggtatt gtatatccca ctgcactgag agggctcaat cagctggatt   4151

tgagttctgg aacacacatc atccccaccc ctccccccagc gtgggctcac cattcttagt   4211

cctttctcaa gtgggacctt caactttctg tgaacaccca gtctgcgtcc tgggtctgct   4271

aggttcgatg atggcgaact cgtatctgca tccggtgcaa gttttagctg gcagaggtga   4331

gaccggtggt gctggtctgc ctttgccaac tatagccagt ctggagactt gataaaatac   4391

ttcagtgaga ccagcttctc atcaacttgg gcccggcgtg ctgggcctga aagtcacact   4451

acatgcactg cctttgggag tcagctcact ccctgctccc acctggaacc ttgccagcgt   4511

gaaggaggct tccaggtact tcaccctgtc aaccacctct gaatccccac caggcgcctt   4571

cctgggtgga ttcaacaaga tgattttgcc ctttcccagt tctctccttc actttggcat   4631

cagttgtttt ctatgaaaac agtggattgg ttgggttttg tgcagggtct tgggttagag   4691

ccaaaatgga tttgaggatg agtatttttt tttttggttt tgtatatttt gtacattaat   4751

aataaacagt ggaaagagaa aaaaaaaaaa                                    4781
```

<210> 29
<211> 968
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (184)..(411)

<220>
<221> misc_feature
<222> (1)...(968)
<223> n = a,t,c or g

<400> 29
```
ttttttttnt attattatac tttaagtttt agggtacatg tgcacaacgt gcaggtttgt     60

tacatatgta tacatgtgcc atgttagtgt gctgcaccca ttaactcgtc atttagcatt    120

aggtatatct cctacagcta tccctccccc cttcccccac cccacaaaag gtcccagtgt    180

gtg  atg ttc ccc ttc ctg tgt cca tgt gtt ctc att gtt caa ttc cca     228
     Met Phe Pro Phe Leu Cys Pro Cys Val Leu Ile Val Gln Phe Pro
       1               5                  10                  15

cct atg agt gag aac atg tgg gtt tgg ttt ttt tgt cct tgc gat agt     276
Pro Met Ser Glu Asn Met Trp Val Trp Phe Phe Cys Pro Cys Asp Ser
                 20                  25                  30

ttg ctg aga atg atg gtt tcc agc ttc atc cat gtc cct gcg aag gac     324
Leu Leu Arg Met Met Val Ser Ser Phe Ile His Val Pro Ala Lys Asp
             35                  40                  45

atg aac tca ccc ttt ttt atg gaa tac tac aca gcc ata aaa agg aat     372
Met Asn Ser Pro Phe Phe Met Glu Tyr Tyr Thr Ala Ile Lys Arg Asn
         50                  55                  60

gac aac aca tcc ctt gca ggg aca tgg atg gag caa tag gccattatcc     421
Asp Asn Thr Ser Leu Ala Gly Thr Trp Met Glu Gln  *
     65                  70                  75

gtagcaaact aatgcaggaa gagaaaaccc agtactgcat gttctcactt ataagtggga    481

gctaaatgat gagaacacat gaacacaaag aggggaacag acactagggc cgtttagaag    541

ttggcgggtg gtttgctttt ttntttagnt acangattta ttagnaatgg gtactaggct    601

gaataccgtt gtggatggta gtaatcgggt gaacaaagcg cccatgttca caattttagc    661

ttatttttc tggttttcgt catggtaccc cttgtgagtg tgagttatat gtttggttga    721

tatttatcgg ttattggatt tggtccctgg gtggtgtttt cgtggagtgg tggggttgtt    781

gtgattcttg tgtttggctg tggtgtttaa aagaggtgta ggttggtgac ccgggtgttg    841

tccacgttat tggctttgat gaggcccccc tttgtgggtg gtgtttggtg ttgttttggt    901
```

```
ctggtcggtg gccctatttg tgctttttgt tggtagtggc ttcctggtgg tcggtgagcg    961

gtgtgtt                                                              968
```

<210> 30
<211> 1082
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (168)..(725)

<400> 30

```
ttttgaaaaa ctgggatagc agtacccagc tggctagcgt ttaaacttaa gcttggtacc     60

gagctcggat ccactagtcc agtgtggtgg aattccggcc aagtagagct ccgtcctgac    120

gcgccgcctc ccgtgggctc cggccggcta agccgcggcg gacaact  atg ctg aaa     176
                                                      Met Leu Lys
                                                        1

gcc aag atc ctc ttc gtg ggg cct tgc gag agt gga aaa act gtt ttg      224
Ala Lys Ile Leu Phe Val Gly Pro Cys Glu Ser Gly Lys Thr Val Leu
  5                  10                  15

gcc aac ttt ctg aca gaa tct tct gac atc act gaa tac agc cca acc      272
Ala Asn Phe Leu Thr Glu Ser Ser Asp Ile Thr Glu Tyr Ser Pro Thr
 20                  25                  30                  35

caa gga gtg agg atc cta gaa ttt gag aac ccg cat gtt acc agc aac      320
Gln Gly Val Arg Ile Leu Glu Phe Glu Asn Pro His Val Thr Ser Asn
                 40                  45                  50

aac aaa ggc acg ggc tgt gaa ttc gag cta tgg gac tgt ggt ggc gat      368
Asn Lys Gly Thr Gly Cys Glu Phe Glu Leu Trp Asp Cys Gly Gly Asp
            55                  60                  65

gct aag ttt gag tcc tgc tgg ccg gcc ctg atg aag gat gct cat gga      416
Ala Lys Phe Glu Ser Cys Trp Pro Ala Leu Met Lys Asp Ala His Gly
         70                  75                  80

gtg gtg atc gtc ttc aat gct gac atc cca agc cac cgg aag gaa atg      464
Val Val Ile Val Phe Asn Ala Asp Ile Pro Ser His Arg Lys Glu Met
     85                  90                  95

gag atg tgg tat tcc tgc ttt gtc caa cag ccg tcc tta cag gac aca      512
Glu Met Trp Tyr Ser Cys Phe Val Gln Gln Pro Ser Leu Gln Asp Thr
100                 105                 110                 115

cag tgt atg cta att gca cac cac aaa cca ggc tct gga gat gat aaa      560
Gln Cys Met Leu Ile Ala His His Lys Pro Gly Ser Gly Asp Asp Lys
                120                 125                 130
```

```
gga agc ctg tct ttg tcg cca ccc ttg aac aag ctg aag ctg gtg cac      608
Gly Ser Leu Ser Leu Ser Pro Pro Leu Asn Lys Leu Lys Leu Val His
            135                 140                 145

tca aac ctg gaa gat gac cct gag gag atc cgg atg gaa ttc ata aag      656
Ser Asn Leu Glu Asp Asp Pro Glu Glu Ile Arg Met Glu Phe Ile Lys
        150                 155                 160

tat tta aaa agc ata atc aac tcc atg tct gag agc aga gac agg gag      704
Tyr Leu Lys Ser Ile Ile Asn Ser Met Ser Glu Ser Arg Asp Arg Glu
    165                 170                 175

gag atg tca att atg acc tag cc agccttcacc tgggactgcc acatccccag      757
Glu Met Ser Ile Met Thr  *
180                 185

tgaaatcagc atgtttctcg gtgcagatct gaaatcacat ccagctcctg atgttttctt    817

ctccctctga ctgcagagga agtgttccta cctgcaggaa ggcacctgtc acacagggcg    877

ttcactcaga ccatctgtgc tctgccctga gttcagttga gaaaatccta ttatcaaatt    937

tggatttcct ggccccagaa cttcccaaag acctgtaaaa tggagggatt taccacctca    997

catatgtcca gttaaacagt ttgtggactt gtaaccgtcg cagcccaatg atacaacagt   1057

agtttaatca cgtgaaaaaa aaaaa                                        1082
```

<210> 31
<211> 1517
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (666)..(1406)

<220>
<221> misc_feature
<222> (1)...(1517)
<223> n = a,t,c or g

<400> 31

```
ctcgaaatcg atactttgcc ggaccggnnc ggnnnnccgg ggtcgacggg aggcaggagg      60

gccgacccag gggtgctggc cgccctctgt gagaaaactg acaatgacat ccgggcctgc     120

atcaacaccc tgcagttcct gtacagccgg ggccagcggg agctgagcgt gcgggacgtg     180

caggccacac gcgtgggcct caaggaccag cgcagagggc tcttctcggt gtggcaggag     240

gtcttccagc tgcctcgagc ccagagcacc ccacctgcag gcgccgtgtg ggccaggacc     300

ccgccctgcc tgctgacaca ctcctgctgg gtgacgggga cgcgggctcc ctcacctccg     360
```

```
cctcacagcg attctaccgt gtcctgcatg ccgctgcctc tgcgggcgag cacgagaagg    420

tggtccaggg cttgtttgac aacttcctgc gtctgcggct gcgagactcc agcctgggtg    480

ctgtgtgtgt ggccctcgac tggctggcct tcgatgacct gctggcgggg gctgctcatc    540

acagccagag cttccagctg ctgcgctacc cacccttcct gcccgtggcc ttccatgtgc    600

tgtttgcttc cagccacaca cccaggatca ccttccccag cagccagcag gaggcccaga    660

accgg    atg agc cag atg agg aac ctg atc cag acg ctg gtg tcc ggc      707
         Met Ser Gln Met Arg Asn Leu Ile Gln Thr Leu Val Ser Gly
           1               5                  10

atc gcg cca gcc acg cgc agc cgg gcc acg ccc cag gcc ctg ctc ctc      755
Ile Ala Pro Ala Thr Arg Ser Arg Ala Thr Pro Gln Ala Leu Leu Leu
 15                  20                  25                  30

gat gcc ctc tgc ctg ctc ctg gac att ctt gcg ccc aag ctc cgc ccc      803
Asp Ala Leu Cys Leu Leu Leu Asp Ile Leu Ala Pro Lys Leu Arg Pro
                35                  40                  45

gtg agc aca cag ctg tac agc acc cgt gaa aag caa cag ctg gcc agc      851
Val Ser Thr Gln Leu Tyr Ser Thr Arg Glu Lys Gln Gln Leu Ala Ser
            50                  55                  60

ctg gtg ggc acg atg ctc gct tac agc ctg acc tac cgc cag gag cgc      899
Leu Val Gly Thr Met Leu Ala Tyr Ser Leu Thr Tyr Arg Gln Glu Arg
         65                  70                  75

acg ccc gat ggc cag tac atc tac agg ctg gag ccg aac gtg gag gaa      947
Thr Pro Asp Gly Gln Tyr Ile Tyr Arg Leu Glu Pro Asn Val Glu Glu
     80                  85                  90

ctc tgc cgc ttc cct gag ctg cct gcc cgc aag ccc ctc acc tac cag      995
Leu Cys Arg Phe Pro Glu Leu Pro Ala Arg Lys Pro Leu Thr Tyr Gln
 95                  100                 105                 110

acg aag cag ctc atc gcc cgc gag atc gag gtg gag aag atg cgg cgg     1043
Thr Lys Gln Leu Ile Ala Arg Glu Ile Glu Val Glu Lys Met Arg Arg
                115                 120                 125

gcg gag gct tct gcc cgg gta gag aac agc ccc cag gtg gat ggg agc     1091
Ala Glu Ala Ser Ala Arg Val Glu Asn Ser Pro Gln Val Asp Gly Ser
            130                 135                 140

ccc cca ggg ctc gag ggt ctg ctg ggg ggc att ggg gag aaa ggg gtg     1139
Pro Pro Gly Leu Glu Gly Leu Leu Gly Gly Ile Gly Glu Lys Gly Val
        145                 150                 155

cac cga cct gcc cca cgc aac cat gag cag cgg ctg gag cac atc atg     1187
His Arg Pro Ala Pro Arg Asn His Glu Gln Arg Leu Glu His Ile Met
    160                 165                 170

agg cga gcg gcc cgg gag gaa cag cct gag aag gac ttc ttt gga cgt     1235
Arg Arg Ala Ala Arg Glu Glu Gln Pro Glu Lys Asp Phe Phe Gly Arg
```

```
175                 180                 185                 190

gtg gtc gtc agg agc aca gca gtc ccg agt gca ggg gac acg gcc ccg     1283
Val Val Val Arg Ser Thr Ala Val Pro Ser Ala Gly Asp Thr Ala Pro
                195                 200                 205

gag cag gac tca gtg gag cgg cgc atg ggc aca gcg gtg ggc agg agc     1331
Glu Gln Asp Ser Val Glu Arg Arg Met Gly Thr Ala Val Gly Arg Ser
            210                 215                 220

gag gtc tgg ttc cgc ttc aac gag ggt gtc tcc aac gcc gtg cgg cgc     1379
Glu Val Trp Phe Arg Phe Asn Glu Gly Val Ser Asn Ala Val Arg Arg
        225                 230                 235

agc ctg tac atc agg gac ttg ctc tag ttctc tgagccgcgg acatgccctc     1431
Ser Leu Tyr Ile Arg Asp Leu Leu  *
    240                 245

gcattgcttc ccgcagagtg cagagacagg aagctggaga tgtctttata aagtcacacc   1491

tttacagact gtaaaaaaaa aaaaaa                                       1517
```

<210> 32
<211> 618
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (498)..(572)

<400> 32

```
ttttgcctaa gatacaatga taagtaaaaa catgttacag agtaataacg tgttgcatag     60

gacaatactg taaatgtggt tcattccatc aacagacact gagccctgcg acgtgcctgg    120

ctctattcta catctgaggg acacaaggtg aacaagacca ggccactgta ttcaacatct    180

acatttaatg gaaatttttg aaagaagact tgagagatta taacagtggt cttcagcatc    240

agggagtagg cctagaagaa gaggaggtca agaagtggct tttcttatta tcttcttaac    300

tcttcaaatt tttactatga gcaattatta ttttttatta aaattttagg ccaggcttat    360

ggctgtaatc ctagcagttt gggaggccaa ggtgagcgga tcacttgagg ttgggagttc    420

gagaccagcc tgaccaacat ggagaaactc tgtctctact taaaaaaaaa tacaaaatta    480

gccaggcatg gtggcac   atg cct gta gtc cca gct act cag gag act gag      530
                    Met Pro Val Val Pro Ala Thr Gln Glu Thr Glu
                      1               5                  10

gca ggg gaa ttg cct gaa cct ggg aga cag agg ttg caa tga gccaaga      579
Ala Gly Glu Leu Pro Glu Pro Gly Arg Gln Arg Leu Gln  *
```

```
             15                  20                  25

tcacgcctcg tgccgaattc ttggcctcga gggccaaat                          618
```

<210> 33
<211> 1188
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (313)..(849)

<400> 33

```
aaggatcctt aattaaatta atccccccc cccggggaga aacgttctca ctcgctctct     60

gctcgctgcg ggcgctcccc gccctctgct gccagaacct tggggatgtg cctagacccg    120

gcgcagcaca cgtccgggcc aaccgcgagc agaacaaacc tttggcgggc ggccaggagg    180

ctcctcccca gccaccgccc ccctccagcg cctttttttc cccccataca atacaagatc    240

ttccttcctc agttccctta aagcacagcc cagggaaacc tcctcacagt tttcatccag    300

ccacgggcca gc    atg tct ggg ggc aaa tac gta gac tcg gag gga cat      348
                 Met Ser Gly Gly Lys Tyr Val Asp Ser Glu Gly His
                   1               5                  10

ctc tac acc gtt ccc atc cgg gaa cag ggc aac atc tac aag ccc aac      396
Leu Tyr Thr Val Pro Ile Arg Glu Gln Gly Asn Ile Tyr Lys Pro Asn
         15                  20                  25

aac aag gcc atg gca gac gag ctg agc gag aag caa gtg tac gac gcg      444
Asn Lys Ala Met Ala Asp Glu Leu Ser Glu Lys Gln Val Tyr Asp Ala
     30                  35                  40

cac acc aag gag atc gac ctg gtc aac cgc gac cct aaa cac ctc aac      492
His Thr Lys Glu Ile Asp Leu Val Asn Arg Asp Pro Lys His Leu Asn
 45                  50                  55                  60

gat gac gtg gtc aag att gac ttt gaa gat gtg att gca gaa cca gaa      540
Asp Asp Val Val Lys Ile Asp Phe Glu Asp Val Ile Ala Glu Pro Glu
                 65                  70                  75

ggg aca cac agt ttt gac ggc att tgg aag gcc agc ttc acc acc ttc      588
Gly Thr His Ser Phe Asp Gly Ile Trp Lys Ala Ser Phe Thr Thr Phe
             80                  85                  90

act gtg acg aaa tac tgg ttt tac cgc ttg ctg tct gcc ctc ttt ggc      636
Thr Val Thr Lys Tyr Trp Phe Tyr Arg Leu Leu Ser Ala Leu Phe Gly
         95                  100                 105

atc ccg atg gca ctc atc tgg ggc att tac ttc gcc att ctc tct ttc      684
Ile Pro Met Ala Leu Ile Trp Gly Ile Tyr Phe Ala Ile Leu Ser Phe
```

```
    110                 115                 120

ctg cac atc tgg gca gtt gta cca tgc att aag agc ttc ctg att gag      732
Leu His Ile Trp Ala Val Val Pro Cys Ile Lys Ser Phe Leu Ile Glu
125                 130                 135                 140

att cag tgc atc agc cgt gtc tat tcc atc tac gtc cac acc gtc tgt      780
Ile Gln Cys Ile Ser Arg Val Tyr Ser Ile Tyr Val His Thr Val Cys
                145                 150                 155

gac cca ctc ttt gaa gct gtt ggg aaa ata ttc agc aat gtc cgc atc      828
Asp Pro Leu Phe Glu Ala Val Gly Lys Ile Phe Ser Asn Val Arg Ile
            160                 165                 170

aac ttg cag aaa gaa ata taa at gacatttcaa ggatagaagt atacctgatt      881
Asn Leu Gln Lys Glu Ile  *
        175

ttttttcctt ttaattttcc tggtgccaat ttcaagttcc aagttgctaa tacagcaaca    941

atttatgaat tgaattatct tggttgaaaa taaaaagatc actttctcag ttttcataag   1001

tattatgtct cttctgagct atttcatcta tttttggcag tctgaatttt taaaacccat   1061

ttaaattttt ttccttacct ttttatttgc atgtggatca accatcgctt tattggctga   1121

gatatgaaca tattgttgaa aggtaatttg agagaaatat gaagaactga ggaggaaaaa   1181

aaaaaaa                                                            1188
```

<210> 34
<211> 920
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (515)..(697)

<400> 34

```
taagcttgcg gccgcaattt ttttttttt ttttttgtat tttttggtag agacgggatt     60

tcactatgtt ggtcaggctg gtctcgaact cccgaccgca agtgatccac ccgccttggc    120

ctcccaaagt gctgggatta caagcttgag ccactgcacc cagcctggaa agtatattta    180

tgaaaggttt gcactccaca aaagcatctt tgctagggtg tcaaggaaga gatcactaaa    240

ccaaccccaa cacatccata caattccagc aatctagaga gggctggtcc ttttcctttt    300

ctggattatt ttctgttctc agtaaaacaa gtatttactg tgatactgaa acactgggaa    360

attaacactg attaagatat tttaaacact gagtcttaat tataacagaa ccagttttca    420
```

```
tcagaatgct tttacgtcac attcagtgaa gtgttacgct aatatattct acagccctga    480

agatagaaaa aaggtttctc tccaggtatg agat    atg gta caa aaa tac att     532
                                         Met Val Gln Lys Tyr Ile
                                           1               5

ttt cca cat aca aaa gag aga aaa aaa caa aga cat gtg gcg ggt ggc     580
Phe Pro His Thr Lys Glu Arg Lys Lys Gln Arg His Val Ala Gly Gly
            10                  15                  20

gag ggg agg ccc aat ccc aac acc cta caa ggt tcc atg gaa tgg aga     628
Glu Gly Arg Pro Asn Pro Asn Thr Leu Gln Gly Ser Met Glu Trp Arg
        25                  30                  35

agg aac aaa aaa atc ccc aat tat ttt ggg gta aga tgt gcc cca gaa     676
Arg Asn Lys Lys Ile Pro Asn Tyr Phe Gly Val Arg Cys Ala Pro Glu
    40                  45                  50

aag gtg aaa tct atg caa taa aa cccaggtttt cttcaaatct agcatctagg     729
Lys Val Lys Ser Met Gln  *
 55                  60

atttctatca gagtttcaaa taatcagaat ttctatcaga atttctaccc tgaggtgaca    789

cctactaact gtaggttctt tcattaaaaa tgaagacatc tttcaccaga atgtatcaag    849

ctataaaact ggcttcagag cctacactta gccagagtgg aaaaaaaaaa tagtgcatat    909

tttcgacagc a                                                        920
```

<210> 35
<211> 1233
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (133)..(1122)

<400> 35

```
gtgatcatcg acgcctgcgg taccggtccg gaattcccgg gtcgacccac gcgtccgggc     60

ggcttcctag tgagtcggcg gctgatttag aaggaggttc aggctacggt gagccgaagc    120

cacacaggag cc    atg gaa gtg gca gag ccc agc agc ccc act gaa gag     168
                 Met Glu Val Ala Glu Pro Ser Ser Pro Thr Glu Glu
                   1               5                  10

gag gag gag gaa gag gag cac tcg gca gag cct cgg ccc cgc act cgc     216
Glu Glu Glu Glu Glu Glu His Ser Ala Glu Pro Arg Pro Arg Thr Arg
        15                  20                  25

tcc aat cct gaa ggg gct gag gac cgg gca gta ggg gca cag gcc agc     264
Ser Asn Pro Glu Gly Ala Glu Asp Arg Ala Val Gly Ala Gln Ala Ser
```

```
     30                  35                  40

gtg ggc agc cgc agc gag ggt gag ggt gag gcc gcc agt gct gat gat      312
Val Gly Ser Arg Ser Glu Gly Glu Gly Glu Ala Ala Ser Ala Asp Asp
 45                  50                  55                  60

ggg agc ctc aac act tca gga gcc ggc cct aag tcc tgg cag gtg ccc      360
Gly Ser Leu Asn Thr Ser Gly Ala Gly Pro Lys Ser Trp Gln Val Pro
                 65                  70                  75

ccg cca gcc cct gag gtc caa att cgg aca cca agg gtc aac tgt cca      408
Pro Pro Ala Pro Glu Val Gln Ile Arg Thr Pro Arg Val Asn Cys Pro
             80                  85                  90

gag aaa gtg att atc tgc ctg gac ctg tca gag gaa atg tca ctg cca      456
Glu Lys Val Ile Ile Cys Leu Asp Leu Ser Glu Glu Met Ser Leu Pro
         95                 100                 105

aag ctg gag tcg ttc aac ggc tcc aaa acc aac gcc ctc aat gtc tcc      504
Lys Leu Glu Ser Phe Asn Gly Ser Lys Thr Asn Ala Leu Asn Val Ser
    110                 115                 120

cag aag atg att gag atg ttc gtg cgg aca aaa cac aag atc gac aaa      552
Gln Lys Met Ile Glu Met Phe Val Arg Thr Lys His Lys Ile Asp Lys
125                 130                 135                 140

agc cac gag ttt gca ctg gtg gtg gtg aac gat gac acg gcc tgg ctg      600
Ser His Glu Phe Ala Leu Val Val Val Asn Asp Asp Thr Ala Trp Leu
                145                 150                 155

tct ggc ctg acc tcc gac ccc cgc gag ctc tgt agc tgc ctc tat gat      648
Ser Gly Leu Thr Ser Asp Pro Arg Glu Leu Cys Ser Cys Leu Tyr Asp
            160                 165                 170

ctg gag acg gcc tcc tgt tcc acc ttc aat ctg gaa gga ctt ttc agc      696
Leu Glu Thr Ala Ser Cys Ser Thr Phe Asn Leu Glu Gly Leu Phe Ser
        175                 180                 185

ctc atc cag cag aaa act gag ctt ccg gtc aca gag aac gtg cag acg      744
Leu Ile Gln Gln Lys Thr Glu Leu Pro Val Thr Glu Asn Val Gln Thr
    190                 195                 200

att ccc ccg cca tat gtg gtc cgc acc atc ctt gtc tac agc cgt cca      792
Ile Pro Pro Pro Tyr Val Val Arg Thr Ile Leu Val Tyr Ser Arg Pro
205                 210                 215                 220

cct tgc cag ccc cag ttc tcc ttg acg gag ccc atg aag aaa atg ttc      840
Pro Cys Gln Pro Gln Phe Ser Leu Thr Glu Pro Met Lys Lys Met Phe
                225                 230                 235

cag tgc cca tat ttc ttc ttt gac gtt gtt tac atc cac aat ggc act      888
Gln Cys Pro Tyr Phe Phe Phe Asp Val Val Tyr Ile His Asn Gly Thr
            240                 245                 250

gag gag aag gag gag gag atg agt tgg aag gat atg ttt gcc ttc atg      936
Glu Glu Lys Glu Glu Glu Met Ser Trp Lys Asp Met Phe Ala Phe Met
        255                 260                 265
```

```
ggc agc ctg gat acc aag ggt acc agc tac aag tat gag gtg gca ctg       984
Gly Ser Leu Asp Thr Lys Gly Thr Ser Tyr Lys Tyr Glu Val Ala Leu
    270                 275                 280

gct ggg cca gcc ctg gag ttg cac aac tgc atg gcg aaa ctg ttg gcc      1032
Ala Gly Pro Ala Leu Glu Leu His Asn Cys Met Ala Lys Leu Leu Ala
285                 290                 295                 300

cac ccc ctg cag cgg cct tgc cag agc cat gct tcc tac agc ctg ctg      1080
His Pro Leu Gln Arg Pro Cys Gln Ser His Ala Ser Tyr Ser Leu Leu
                305                 310                 315

gag gag gag gat gaa gcc att gag gtt gag gcc act gtc tga accatcc      1129
Glu Glu Glu Asp Glu Ala Ile Glu Val Glu Ala Thr Val  *
            320                 325                 330

ctgtacatct gcaccttctt gtgcaaggaa gtccttggcc taaagccttg gttctcaaac    1189

tgggttcctt gggacctccg gggtgggggg gttccaggag gcat                     1233
```